

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, 15/63, 15/00, C12P 21/02, C07K 14/47, 16/00, G01N 33/53, C12Q 1/68		A1	(11) International Publication Number: WO 99/15660 (43) International Publication Date: 1 April 1999 (01.04.99)
(21) International Application Number: PCT/US98/19979 (22) International Filing Date: 24 September 1998 (24.09.98)		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>	
(30) Priority Data: 60/059,863 24 September 1997 (24.09.97) US			
(71) Applicant (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): LIU, Qingyun [CN/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). BAILY, Wendy, J. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). MCDONALD, Terrence, P. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).			
(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).			
(54) Title: G-PROTEIN COUPLED GLYCOPROTEIN HORMONE RECEPTOR HG38			
(57) Abstract <p>This invention provides a novel G-protein coupled glycoprotein hormone receptor HG38, mutant and polymorphic forms of the receptor, nucleic acids encoding the same, expression vectors including the nucleic acids, host cells transformed with nucleic acids, transgenic knockout animals lacking the receptor and transgenic animals expressing a non-native receptor gene, antibodies against the receptor and polypeptides thereof and assays for modulators, agonists and antagonists of the receptor. The receptor proteins and polypeptides, nucleic acids, cells, animals and assays of this invention are useful in drug screening and development, diagnosis and therapeutic applications.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TITLE OF THE INVENTION

G-PROTEIN COUPLED GLYCOPROTEIN HORMONE RECEPTOR
HG38

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/059,863, filed 9/24/97, the contents of which are incorporated herein by reference in their entirety.

10 STATEMENT REGARDING FEDERALLY-SPONSORED R&D

Not applicable.

REFERENCE TO MICROFICHE APPENDIX

Not applicable.

15

FIELD OF THE INVENTION

This invention relates to a novel G-protein coupled glycoprotein hormone receptor in substantially purified form, and also to mutant or polymorphic forms of the receptor, recombinant nucleic acids 20 encoding the same, recombinant host cells transformed with the nucleic acids, transgenic knockout animals lacking the receptor, transgenic animals expressing a non-native receptor gene, antibodies against the receptor and polypeptides thereof, and the uses of the receptor, recombinant nucleic acids, recombinant host cells and transgenic 25 animals in drug screening and development, diagnosis and therapeutic applications.

BACKGROUND OF THE INVENTION

The G-protein coupled receptor of the present invention is a 30 member of the glycoprotein hormone receptor family. Only three G-protein coupled glycoprotein hormone receptors have been previously reported: the Follicle Stimulating Hormone (FSH) Receptor (Minegish, *et. al.*, 1991. Biomed. Biochem. Res. Comm. 175:1125-1130; Sprengel, *et. al.*, 1990. Mol. Endocrinol. 4:525-530); the Thyroid Stimulating Hormone 35 (TSH) Receptor (Frazier, *et. al.*, 1990. Mol. Endocrinol. 4:1264-1276; Parmentier, *et. al.*, 1990. Science 246:1620-1622) and the Leuteneizing

Hormone/Placental Chorionic Gonadotropin Hormone (LH/hCG) Receptor (Loosfelt, *et. al.*, 1990. *Science* 245:525-528).

The structure and function of the known glycoprotein hormone receptors has been reviewed (Pearce, *et. al.*, 1995. *Q. J. Med.* 88:3-8; Reichert, *et. al.*, 1991. *Trends in Pharmacol. Sci.* 12:219-203). This

5 group of glycoprotein hormone receptors exhibit a structure of the rhodopsin family G-protein coupled receptors. This class of receptors contains seven transmembrane domains ~~form~~ with three extracellular loops and three intracellular loops.

10 The large ligands, including the glycoprotein hormones, bind the N-terminal domain while smaller peptides, amines and other ligands can bind in a pocket formed by the extracellular loops. Upon binding of an activating ligand a conformational change is believed to occur which activates the associated G-protein. In this activation the 15 cytoplasmic loops, particularly the third loop, and the C-terminal domain of the receptor are believed to interact with the G-protein.

18 The receptor associated G-protein can be associated with several cellular signaling pathways. Most common are the adenylate-cyclase/cAMP pathway, the phospholipase C- β /phosphoinositol 20 pathways and the elevation of intracellular Ca^{2+} . These second messenger pathways mediate the action of the receptor ligand within the cell. They also advantageously can be used to assess the activity of a receptor in assays.

25 Receptor activity can be regulated at the cellular level. Extensive activation of a receptor by agonists can result in phosphorylation of the C-terminus and cytoplasmic loops resulting in a rapid desensitization of the receptor. Further, receptors can be regulated by modulators of transcriptional activity on the receptor gene. cAMP responsive elements have been demonstrated within the promoter 30 regions of some G-protein coupled receptor genes. Again, these aspects of cellular biochemistry can advantageously be used to monitor and assess receptor activity in assays, *e.g.*, by monitoring receptor phosphorylation as an indication of the presence of an agonist of the receptor or monitoring transcriptional activity as an indication of the 35 presence of a modulator of receptor gene expression.

Mutations in the known G-protein coupled glycoprotein receptors can lead to or indicate a disease state (Pearce, *et. al.*, 1995). Given the importance of glycoprotein hormone receptors in the endocrine system, HG38 is expected to play an important role in the 5 development and function of skeletal muscle, spinal cord, placenta, and, to a lesser extent, the brain.

SUMMARY OF THE INVENTION

A preferred aspect of the present invention is disclosed in 10 FIGS. 1A-1C and SEQ ID NO:1, a human cDNA encoding a G-protein coupled glycoprotein hormone receptor protein, HG38.

Aspects of this invention are isolated nucleic acid 15 fragments of the HG38 G-protein coupled glycoprotein hormone receptor (SEQ ID NO:1) which encodes mRNA expressing a biologically active novel human receptor. Any such nucleic acid fragment will encode 20 either a protein or protein fragment comprising at least an intracellular G-protein associating domain and/or extracellular ligand binding domain, domains conserved throughout the G-coupled glycoprotein hormone receptor family which exist in the amino acid sequence of HG38 (SEQ ID NO:2). Any such polynucleotide includes but is not necessarily limited to nucleotide substitutions, deletions, additions, 25 amino-terminal truncations and carboxy-terminal truncations such that these mutations encode mRNA which express a protein or protein fragment of diagnostic, therapeutic or prophylactic use, or would be useful for screening for modulators of expression, agonists and/or antagonists of HG38 function.

In particular embodiments, the isolated nucleic acid 30 molecule of the present invention can be a deoxyribonucleic acid molecule (DNA), such as genomic DNA and complementary DNA (cDNA), which can be single (coding or noncoding strand) or double stranded, as well as synthetic DNA, such as a synthesized, single stranded polynucleotide. The isolated nucleic acid molecule of the present invention can also be a ribonucleic acid molecule (RNA). In particular embodiments, the nucleic acid can include the entire 35 sequence of SEQ ID NO:1, a sequence encoding the open reading frame of SEQ ID NO:1, or smaller sequences useful for expressing peptides, or

polypeptides of HG38 protein. In particular embodiments the nucleic acid can have natural, non-natural or modified nucleotides or internucleotide linkages or mixtures of these.

Aspects of the present invention include nucleotide probes

5 and primers derived from the nucleotide and polypeptide sequences disclosed herein as FIGS. 1A-1C, 2, 3A-3E and 4 and SEQ ID NOS: 1 and 2. In particular embodiments of the invention, probes and primers are used to identify or isolate polynucleotides encoding HG38 of FIG. 2 or mutant or polymorphic forms of the HG38 receptor protein or gene.

10 Probe and primers can be highly specific for HG38 nucleotide sequences.

An aspect of this invention is a substantially purified form of the novel G-protein coupled glycoprotein hormone receptor protein, HG38, which is disclosed in FIG. 2 and as set forth in SEQ ID NO:2.

Aspects of the present invention include biologically active

15 fragments and/or mutants of an HG38 protein as set forth as SEQ ID NO:2, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these mutations provide for proteins or protein fragments of diagnostic, therapeutic or prophylactic use and would be

20 useful for screening for modulators, agonists and/or antagonists of HG38 function. In a preferred embodiment, the fragment is a soluble N-terminal fragment that can compete with the receptor for receptor ligands.

Aspects of the present invention include recombinant

25 vectors and recombinant hosts which contain the nucleic acid molecules disclosed throughout this specification. In particular embodiments, the vectors and hosts can be prokaryotic or eukaryotic. In particular embodiments the hosts express HG38 peptides, polypeptides, proteins or fusion proteins. In further embodiments the host cells are used as a

30 source of expression products.

Aspects of the invention are polyclonal and monoclonal antibodies raised in response to either the entire human form of HG38 disclosed herein, or only a fragment, or a single epitope thereof. In a preferred embodiment antibodies are raised against epitopes within the

35 NH₂-terminal domain of HG38. In another preferred embodiment, antibodies are raised to epitopes that are unique to the HG38 receptor.

An Aspect of this invention is the use of the DNA molecules, RNA molecules, recombinant protein and antibodies of the present invention to screen and measure levels of human HG38. The recombinant proteins, DNA molecules, RNA molecules and antibodies 5 lend themselves to the formulation of kits suitable for the detection and typing of human HG38.

Aspects of this invention are assays to detect agonists and antagonists of the HG38 receptor and modulators of the expression of HG38. In particular embodiments of this aspect, cells comprising HG38 10 are used in screening assays including the melanophore system, yeast expressing mammalian adenylate cyclase, yeast pheromone protein surrogate screening, phospholipase second signal screening and the yeast two-hybrid system, all of which are well known and simply adapted by one of skill in the art.

15 An aspect of this invention is tissue typing using probes or antibodies of this invention. In a particular embodiment, polynucleotide probes are used to identify tissues expressing HG38 RNA. In another embodiment, probes or antibodies can be used to identify a type of tissue based on HG38 expression or display of HG38 receptors on the surface of 20 one or more cells.

An aspect of this invention is isolated nucleic acid molecules which are fusion constructions expressing fusion proteins useful in assays to identify compounds which are modulators, agonist or antagonists of wild-type human HG38 activity. A preferred embodiment 25 of this aspect of the invention includes, but is not limited to, glutathione S-transferase GST-HG38 fusion constructs. These fusion constructs include, but are not limited to, all or a portion of the ligand-binding domain of HG38, as an in-frame fusion at the carboxy terminus of the GST gene. The fusion protein is useful to isolate or identify ligands of 30 the HG38 receptor. The disclosure of SEQ ID NOS:1-2 allow the artisan of ordinary skill to construct any such nucleic acid molecule encoding a GST-G-protein coupled glycoprotein hormone receptor fusion protein. Soluble recombinant GST-G-protein coupled glycoprotein hormone receptor fusion proteins can be expressed in various expression systems, 35 including *Spodoptera frugiperda* (Sf21) insect cells (Invitrogen) using a

baculovirus expression vector (e.g., Bac-N-Blue DNA from Invitrogen or pAcG2T from Pharmingen).

An aspect of this invention is pharmaceutical compositions including the HG38 protein, fragments thereof, agonists, antagonists or modulators of HG38 or HG38 polynucleotides.

An aspect of this invention is using polynucleotides according to the invention in methods of gene therapy, for instance in treatment of individuals with the aim of preventing or curing (wholly or partially) disease states associated with mutations in the HG38 gene.

10 This may ease one or more symptoms of the disease. Introduction of nucleic acid may take place in vivo by way of gene therapy vectors and methods

An aspect of this invention is a transgenic animal useful for the study of the tissue and temporal specific expression or activity of the HG38 receptor in a non-human animal. The animal is also useful for studying the ability of a variety of compounds to act as modulators of HG38 receptor activity or expression *in vivo* or, by providing cells for culture or assays, *in vitro*. In an embodiment of this aspect of the invention, the animal is used in a method for the preparation of a further animal which lacks a functional endogenous HG38 gene. In another embodiment, the animal of this aspect is used in a method to prepare an animal which expresses a non-native HG38 gene in the absence of the expression of a endogenous gene. In particular embodiments the non-human animal is a mouse. In further 25 embodiments the non-native HG38 gene is a wild-type human gene or a mutant human HG38 gene.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1C. Schematically depicts the nucleotide sequence of a cDNA polynucleotide encoding the HG38 receptor (SEQ ID NO:1).

FIG. 2. Schematically depicts the full length amino acid sequence of the HG38 receptor protein (SEQ ID NO:2) in single letter code.

35 FIGS. 3A-3E. Schematically depicts the nucleotide sequence of a polynucleotide encoding HG38 (nucleotides 3-3300 of SEQ

ID NO:1) and the translation of the HG38 open reading frame (SEQ ID NO:2).

FIG. 4. Depicts six predicted signal peptide cleavage sites of the HG38 protein. The six sequences depicted are amino acids 9-51, 12-

5 54, 28-70, 13-55, 11-53 and 8-50 of SEQ ID NO:2 respectively, in single letter code.

FIG. 5. Depicts a Multi-tissue Northern blot analysis of the expression of the HG38 receptor gene.

10 DETAILED DESCRIPTION OF THE INVENTION

This invention provides polynucleotides and polypeptides of a human G-coupled glycoprotein hormone receptor, referred to herein as HG38. The polynucleotides and polypeptides are used to further provide expression vectors, host cells comprising the vectors, non-15 human animals transgenic for the polynucleotides, knockout animals, probes and primers, antibodies against the receptor and polypeptides thereof, assays for the presence or expression of HG38 and assays for the identification of modulators, agonists and antagonists of the HG38 receptor.

20 The HG38 gene, receptor and agonists, antagonists and modulators thereof can be useful in the treatment of degenerative diseases of the muscles, *e.g.*, sarcopenia. Further uses can include to stimulate the growth or regeneration of skeletal muscle, to increase or decrease muscle metabolism, and in the treatment of obesity and type II 25 diabetes.

Each document mentioned in this specification is hereby incorporated herein by reference in its entirety.

As used herein a "compound" or a "molecule" is an organic or inorganic assembly of atoms of any size, and can include 30 macromolecules, *e.g.*, peptides, polypeptides, whole proteins, and polynucleotides. The terms are used interchangeable herein.

As used herein, a "candidate" is a molecule or compound that may be a modulator, agonist or antagonist of an HG38 receptor.

As used herein an "agonist" is a compound or molecule that 35 interacts with and activates a polypeptide of an HG38 receptor. An activated HG38 receptor polypeptide can stimulate the cleavage of GTP by

a G protein, activate the adenylate cyclase pathway or activate the phospholipase b pathway.

As used herein an "antagonist" is a compound or molecule that interacts with and inhibits or prevents a polypeptide of an HG38 receptor from becoming activated.

As used herein a "modulator" is a compound or molecule that interacts with an aspect of cellular biochemistry to effect an increase or decrease in the amount of a polypeptide of an HG38 receptor present at the surface of a cell, or in the surrounding serum or media.

10 The change in amount of the receptor polypeptide can be mediated by the effect of a modulator on the expression of the receptor, *e.g.*, the transcription, translation, post-translational processing, translocation or folding of the receptor, or by affecting a component(s) of cellular biochemistry that directly or indirectly participates in the expression of

15 the receptor. Alternatively, a modulator can act by accelerating or decelerating the turnover of the receptor either by direct interaction with the receptor or by interacting with another component(s) of cellular biochemistry which directly or indirectly effects the change.

20 Polynucleotides

A preferred aspect of the present invention is disclosed in FIGS. 1 and SEQ ID NO:1, a human cDNA encoding a G-protein coupled glycoprotein hormone receptor, HG38, disclosed as follows:

25 CGCACCGCCA CTGTCGCCGC TGCAGCCAGG GCTGCTCCGA AGGCCGGCGT
GGCGGCAACC GGCACCTCAG TCCCCGCCGC GCTTCTCCTC GCCGCCACG
CCGTGGGGTC AGGAACGCGG CGTCTGGCGC TGCAGACGCC CGCTGAGTTG
CAGAAGCCCA CGGAGCGGGCG CCCGGCGCAGC CACGGCCCGT AGCAGTCCGG
10 TGCTGCTCTC CGCCCGCGTC CGGCTCGTGG CCCCCCTACTT CGGGCACCAT
GGACACCTCCTC CGGCTCGGTG TGCTCCTGTC CTTGCCTGTG CTGCTGCAGC
TGGCGACCGG GGGCAGCTCT CCCAGGTCTG GTGTGTTGCT GAGGGGCTGC
15 CCCACACACT GTCATTGCGA GCCCCGACGGC AGGATGTTGC TCAGGGTGGA
CTGCTCCGAC CTGGGGCTCT CGGAGCTGCC TTCCAACCTC AGCGTCTTCA
CCTCCTACCT AGACCTCAGT ATGAACAACA TCAGTCAGCT GCTCCCGAAT
20 CCCCTGCCCA GTCTCCGCTT CCTGGAGGAG TTACGTCTTG CGGGAAACGC
TCTGACATAC ATTCCCCAAGG GAGCATTACAC TGCCCTTTAC AGTCTAAAG
TTCTTATGCT GCAGAATAAT CAGCTAAGAC ACGTACCCAC AGAAGCTCTG
CAGAATTTCG CAAAGCCTTCA ATCCCTGCCT CTGGATGCTA ACCACATCAG
25 CTATGTGCC CCAAGCTGTT TCAGTGGCCT GCATTCCCTG AGGCACCTGT
GGCTGGATGA CAATGCGTTA ACAGAAATCC CCGTCCAGGC TTTTAGAAGT
30 TTATCGGCAT TGCAAGCCAT GACCTTGGCC CTGAACAAAA TACACCACAT
ACCAGACTAT GCCTTTGGAA ACCTCTCCAG CTTGGTAGTT CTACATCTCC

5 ATAACAATAG AATCCACTCC CTGGGAAAGA AATGCTTGA TGGGCTCCAC
 AGCCTAGAGA CTTAGATTT AAATTACAAT AACCTTGATG AATTCCCCAC
 TGCAATTAGG ACACTCTCCA ACCTTAAAGA ACTAGGATTT CATAGCAACA
 ATATCAGGTC GATACCTGAG AAAGCATTG TAGGCAACCC TTCTCTTATT
 10 ACAATACATT TCTATGACAA TCCCATCCAA TTTGTTGGGA GATCTGCTTT
 TCAACATTAA CCTGAACTAA GAACACTGAC TCTGAATGGT GCCTCACAAA
 TAACTGAATT TCCTGATTAA ACTGGAACCTG CAAACCTGGGA GAGTCTGACT
 TTAACGGAG CACAGATCTC ATCTCTTCT CAAACCGTCT GCAATCAGTT
 ACCTAATCTC CAAGTGCTAG ATCTGTCTTA CAACCTATTAA GAAGATTTAC
 15 CCAGTTTTTC AGTCTGCCAA AAGCTTCAGA AAATTGACCT AAGACATAAT
 GAAATCTACG AAATTAAAGT TGACACTTTC CAGCAGTTGC TTAGCCTCCG
 ATCGCTGAAT TTGGCTTGGAA ACAAAATTGC TATTATTCAC CCCAATGCAT
 TTTCCACTTT GCCATCCCTA ATAAAGCTGG ACCTATCGTC CAACCTCCTG
 TCGTCTTTTC CTATAACTGG GTTACATGGT TTAACTCACT TAAAATTAAC
 20 AGGAAATCAT GCCTTACAGA GCTTGTATTC ATCTGAAAAC TTTCCAGAAC
 TCAAGGTTAT AGAAATGCCT TATGCTTACG AGTGTGTGTC ATTTGGAGTG
 TGTGAGAATG CCTATAAGAT TTCTAATCAA TGGAAATAAG GTGACAAACAG
 CAGTATGGAC GACCTTCATA AGAAAGATGC TGGAAATGTTT CAGGCTCAAG
 ATGAACGTGA CCTTGAAGAT TTCCCTGCTTG ACTTGAGGA AGACCTGAAA
 25 GCCCTTCATT CAGTGCAGTG TTCACCTTCC CCAGGCCCCCT TCAAACCCCTG
 TGAACACCTG CTTGATGGCT GGCTGATCAG AATTGGAGTG TGGACCATAG
 CAGTTCTGGC ACTTACTTGT AATGCTTGG TGACTTCAC AGTTTCAGA
 TCCCCTCTGT ACATTTCCCC CATTAAACTG TTAATTGGGG TCATCGCAGC
 AGTGAACATG CTCACGGGAG TCTCCAGTGC CGTGCTGGCT GGTGTGGATG
 30 CGTTCACTTT TGGCAGCTTT GCACGACATG GTGCCCTGGTG GGAGAATGGG
 GTTGGTTGCC ATGTCAATTGG TTTTTTGTCC ATTGTTGCTT CAGAATCATC
 TGTTTCTCTG CTTACTCTGG CAGCCCTGGA GCGTGGGTTC TCTGTGAAAT
 ATTCTGCAAA ATTTGAAACG AAAGCTCCAT TTTCTAGCCT GAAAGTAATC
 ATTTTGCTCT GTGCCCTGCT GCCCTTGACC ATGCCCGCAG TTCCCTGCT
 35 GGGTGGCAGC AAGTATGGCG CCTCCCCCTCT CTGCCCTGCCT TTGCCCTTTG
 GGGAGCCCAG CACCATGGGC TACATGGTCG CTCTCATCTT GCTCAATTCC
 CTTTGCTTCC TCATGATGAC CATTGCCTAC ACCAAGCTCT ACTGCAATT
 GGACAAGGGA GACCTGGAGA ATATTTGGGA CTGCTCTATG GTAAAACACA
 TTGCCCTGTT GCTCTTCACC AACTGCATCC TAAACTGCCC TGTGGCTTTC
 40 TTGTCCTTCT CCTCTTTAAT AAACCTTACA TTTATCAGTC CTGAAGTAAT
 TAAGTTTATC CTTCTGGTGG TAGTCCCCT TCCCTGCATGT CTCAATCCCC
 TTCTCTACAT CTTGTTCAAT CCTCACTTTA AGGAGGATCT GGTGAGCCTG
 AGAAAGCAAA CCTACGTCTG GACAAGATCA AAACACCCAA GCTTGATGTC
 AATTAACTCT GATGATGTCG AAAAACAGTC CTGTGACTCA ACTCAAGCCT
 45 TGGTAACCTT TACCAAGCTCC ACCATCACCT ATGACCTGCC TCCCAGTTCC
 GTGCCATCAC CAGCTTATCC AGTGAATGAG AGCTGCCATC TTTCCCTCTGT
 GGCATTTGTC CCATGTCTCT AATTAATATG TGAAGGAAAA TGTTTCAAA
 GGTTGAGAAC CTGAAAATGT GAGATTGAGT ATATCAGAGC AGTAATTAAT
 AAGAAGAGCT GAGGTGAAAC TCGGTTAAA AACCAAAAAA GAATCTCTCA
 50 GTTAGTAAGA AAAGGCTGAA AACCTCTTGA TACCTTGAGAG TGAATATAAG
 TCTAAATGCT GCTTTGTATA ATTGTTTCAG GTAAGGGATA GATCGATCAC
 ACTATTTAAG TGAGCCAGA TCAAAAAAGC AGATTGAAAT TTTCTTTAGA
 AAAGATTCTC CATGATTGA ATTGCATTCT CTTTAAACTC ACCAATGTAA
 TCATTTGGG AGGTGGGAGA ACCCCCTTGT TTTCAAATG GGTTTATTAA
 AACCCACAAA CTCAAGAGGT TGTTGGGGAA ATTAGGAAAAA TAAGGGTTTT
 CAATGACCTA CATTGCTAGG TAGAGGCTGT GATCCATGGG TTTCATTCTA
 ATGACCATGT GAGATGTTG GTCTTCCTT GCTTTCTCA GAAAGATCCT
 TCTAAGGCAC AAATCCCTTA GATGGATAAT GTAAGGTATT GTTAACTCAC
 TCATATTGAG ATCATTTTA GAGATACCAAG GTTTTATGTA TCAGCACTAG

5 ATGGTTCCAC CCTCATGGGA TAAAAGTGCT TACAAGTATT TTGAAAGAAA
 AACTGACCAA AATTCTTAAA TTGTTACTAA GGCAATCATG CACAGGTGAC
 GTATGTCTTA TCTGATTGT TTTTAACTCC TTGGTGCCCC AAGCTCAGAA
 GGGAAATTCCA CTGCCAGCAA TGAACATACC TGGAAAAGAA AGTAAGCAAT
 CTGGGATTTT TTTTCTGGGT TAGTAAAGAA TTTTGCAAT AAGTTTATC
 AGTTGATTCA AACTGATGTG CATCTTAATG ATCAAATGTG CACATTACAT
 AAATTAAGTC CACTGATACA ACTTCTTACA CATGTATCTC TAGTAGCTCT
 GGCAAACCCA ATATCTGACA CCACCTTGGGA CTCAAGAGAC TCAGTAACGT
 ATTATCCTGT TTATTCTAGCT TGTTTTAGC TGTGTTCTCT CTGGATAACC
 10 CACTTGATGT TAGGAACATT ATTTCTCTGC TTATTCCATA TTAATACTGT
 GTTAGGTATT TTAAGAAGCA AGTTATTAAGA TAAGAAAAGT CAAAGTATTA
 ATTCTTACCT TCTATTATCC TATATTAGCT TCAATACATC CAAACCAAAT
 GGCTGTTAGG TAGATTATT TTTATATAAG CATGTTTATT TTGATCAGAT
 GTTTTAACCTT GGATTGAAA AAATACATT ATGAGATGTT TTATAAGATG
 15 TGTAATATA GAACTGTATT TATTACTATA GTAAAGGTTG AGTAACATTA
 AGGACCATGA TAATGATAAT AAACCTTGTA CAGTGGCATA TTCTTTGATT
 TATATTGTGT TTCTCTGCC ATTTCTTTA AATTCAATTAA CTGTATATAT
 GTAAATATAT AGTACTTGTA AATAGATTCC AAATTTGCTT TTCTATTGGG
 20 TAAAAAATAA ATTTGTAATA AAATGTGTGA CTATGAAACA AAAAAAA
 AAAAAAA

The isolated nucleic acid molecule of the present invention
 can include a deoxyribonucleic acid molecule (DNA), such as genomic
 DNA and complementary DNA (cDNA), which can be single (coding or
 25 noncoding strand) or double stranded, as well as synthetic DNA, such as
 a synthesized, single stranded polynucleotide. The isolated nucleic acid
 molecule of the present invention can also include a ribonucleic acid
 molecule (RNA).

The present invention also relates to recombinant vectors
 30 and recombinant hosts, both prokaryotic and eukaryotic, which contain
 the substantially purified nucleic acid molecules disclosed throughout
 this specification.

As used herein a "polynucleotide" is a nucleic acid of more
 than one nucleotide. A polynucleotide can be made up of multiple
 35 polynucleotide units that are referred to by description of the unit. For
 example, a polynucleotide can comprise within its bounds a
 polynucleotide(s) having a coding sequence(s), a polynucleotide(s) that is
 a regulatory region(s) and/or other polynucleotide units commonly used
 in the art.

40 An "expression vector" is a polynucleotide having
 regulatory regions operably linked to a coding region such that, when in
 a host cell, the vector can direct the expression of the coding sequence.

The use of expression vectors is well known in the art. Expression vectors can be used in a variety of host cells and, therefore, the regulatory regions are preferably chosen as appropriate for the particular host cell.

5 A "regulatory region" is a polynucleotide that can promote or enhance the initiation or termination of transcription or translation of a coding sequence. A regulatory region includes a sequence that is recognized by the RNA polymerase, ribosome, or associated transcription or translation initiation or termination factors of a host 10 cell. Regulatory regions that direct the initiation of transcription or translation can direct constitutive or inducible expression of a coding sequence.

Polynucleotides of this invention contain full length or partial length sequences of the mammalian HG38 receptor gene.

15 Polynucleotides of this invention can be single or double stranded. If single stranded, the polynucleotides can be a coding, "sense," strand or a complementary, "antisense," strand. Antisense strands can be useful as modulators of the receptor by interacting with RNA encoding the receptor. Antisense strands are preferably less than full length strands 20 having sequences unique or highly specific for RNA encoding the receptor.

The polynucleotides can include deoxyribonucleotides, ribonucleotides or mixtures of both. The polynucleotides can be produced by cells, in cell-free biochemical reactions or through chemical 25 synthesis. Non-natural or modified nucleotides, including inosine, methyl-cytosine, deaza-guanosine, etc., can be present. Natural phosphodiester internucleotide linkages can be appropriate. However, polynucleotides can have non-natural linkages between the nucleotides. Non-natural linkages are well known in the art and include, without 30 limitation, methylphosphonates, phosphorothioates, phosphorodithionates, phosphoroamidites and phosphate ester linkages. Dephospho-linkages are also known, as bridges between nucleotides. Examples of these include siloxane, carbonate, carboxymethyl ester, acetamide, carbamate, and thioether bridges. "Plastic DNA," having, 35 for example, N-vinyl, methacryloxyethyl, methacrylamide or ethyleneimine internucleotide linkages, can be used. "Peptide Nucleic

Acid" (PNA) is also useful and resists degradation by nucleases. These linkages can be mixed in a polynucleotide.

As used herein, "purified" and "isolated" are utilized interchangeably to stand for the proposition that the polynucleotides, 5 proteins and polypeptides, or respective fragments thereof in question has been removed from its *in vivo* environment so that it can be manipulated by the skilled artisan, such as but not limited to sequencing, restriction digestion, site-directed mutagenesis, and subcloning into expression vectors for a nucleic acid fragment as well as 10 obtaining the protein or protein fragment in pure quantities so as to afford the opportunity to generate polyclonal antibodies, monoclonal antibodies, amino acid sequencing, and peptide digestion. Therefore, the nucleic acids claimed herein can be present in whole cells or in cell lysates or in a partially purified or substantially purified form. A 15 polynucleotide is considered purified when it is purified away from environmental contaminants. Thus, a polynucleotide isolated from cells is considered to be substantially purified when purified from cellular components by standard methods while a chemically synthesized nucleic acid sequence is considered to be substantially purified when 20 purified from its chemical precursors.

Polypeptides

The present invention also relates to a substantially purified form of the novel G-protein coupled glycoprotein hormone receptor 25 protein, HG38, which is shown in FIG. 2 and as set forth in SEQ ID NO:2, disclosed as follows in single letter code:

MDTSRLGVLL SLPVLLQLAT GGSSPRSGVL LRGCPTHCHC EPDGRMLLRV
 30 DCSDLGLSEL PSNLSVFTSY LDLSMNNISQ LLPNPLPSLR FLEELRLAGN
 ALTYIPKGAF TGLYSLKVL M LQNNQLRHVP TEALQNLRSQ QSLRLDANHI
 SYVPPSCFSG LHSLRHLWLD DNALTEIPVQ AFRSLSALQA MTLALNKIH
 35 IPDYAFGNLS SLVVLHLHNN RIHSLGKKCF DGLHSLETLD LNYNNLDEF
 TAIRTLSNLK ELGFHSNNIR SIPEKAFVGN PSLITIHFYD NPIQFVGRSA
 FQHLPELRTL TLNGASQITE FPDLTGTANL ESLTLTGAQI SSLPQTVCNQ
 LPNIQVLDLS YNLLEDLPSF SVCQKLQKID LRHNEIYEIK VDTFQQLL
 40 RSLNLAWNKI AIIHPNAFST LPSLIKLDLS SNLSSFPIT GLHGLTHLKL
 TGNHALQSLI SSENFPPELKV IEMPYAYQCC AFGVCENAYK ISNQWNKGDN
 SSMDDLHKKD AGMFQAQDER DLEDFLLDFE EDLKALHSVQ CSPSPGPFKP
 CEHLLDGWLI RIGWWTIAVL ALTCNALVTS TVFRSPLYIS PIKLLIGVIA
 AVNMLTGVS AVLAGVDAFT FGSFARHGA WENGVGCHVI GFLSIFASES
 SVFLLTAA L ERGFSVKYSA KFETKAPFSS LKVIIILLCAL LALTMAAVPL

5 LGGSKYGASP LCLPLPFGEP STMGYMVALI LLNSLCFLMM TIAYTKLYCN
LDKGDLENIW DCSMVKHIAL LLFTNCILNC PVAFLSFSSL INLTFISPEV
IKFILLVVVP LPACLNPLLY ILFNPHFKED LVSLRKQTYV WTRSKHPSLM
SINSDDVEKQ SCDSTQALVT FTSSSITYDL PPSSVPSPAY PVTESCHLSS
VAFVPCL

The present invention also relates to biologically active fragments and mutant or polymorphic forms of HG38 as set forth as SEQ ID NO:2, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these mutations provide for proteins or protein fragments of diagnostic, therapeutic or prophylactic use and would be useful for screening for modulators, agonists and/or antagonists of HG38 function.

10 In a preferred embodiment, the biologically active fragment of HG38 is a soluble N-terminal fragment that can compete with the complete HG38 receptor for ligands of the receptor. Such soluble forms of receptors are well known in the art and can be derived from the polypeptides disclosed herein. It is preferred that soluble N-terminal fragments lack the signal sequence, that is that lack about the first 22 amino acids of SEQ ID NO:2. By "about" it is meant that the fragment need not lack exactly 22 amino acids as it is expected that deletion or removal of more or less can be useful. The important point is not so much the amount deleted but that the N-terminal fragment retains ligand binding activity. Any HG38 fragment can be simply tested for competition with the HG38 receptor using an antagonist assay described herein. The length can also vary. Soluble N-terminal fragments having the sequence of SEQ ID NO:2 up to but not including the seven hydrophobic domains are preferred. For example, it is preferred that soluble N-terminal fragments extend up to about amino acid 557 of SEQ ID NO:2. Again, this need not be an exact endpoint, as other appropriate endpoints can be determined by simple testing, e.g., for binding activity compared to the wild-type.

15 Using the disclosure of polynucleotide and polypeptide sequences provided herein to isolate polynucleotides encoding naturally occurring forms of HG38, one of skill in the art can determine whether such naturally occurring forms are mutant or polymorphic forms of

20

25

30

HG38 by sequence comparison. One can further determine whether the encoded protein, or fragments of any HG38 protein, is biologically active by routine testing of the protein or fragment in a *in vitro* or *in vivo* assay for the biological activity of the HG38 receptor. For example, one can

5 express N-terminal or C-terminal truncations, or internal additions or deletions, in host cells and test for their ability to stimulate the cleavage of GTP by a G protein, activate the adenylate cyclase pathway or activate the phospholipase b pathway.

It is known that there is a substantial amount of

10 redundancy in the various codons which code for specific amino acids. Therefore, this invention is also directed to those DNA sequences encode RNA comprising alternative codons which code for the eventual translation of the identical amino acid, as shown below:

A=Ala=Alanine: codons GCA, GCC, GCG, GCU

15 C=Cys=Cysteine: codons UGC, UGU

D=Asp=Aspartic acid: codons GAC, GAU

E=Glu=Glutamic acid: codons GAA, GAG

F=Phe=Phenylalanine: codons UUC, UUU

G=Gly=Glycine: codons GGA, GGC, GGG, GGU

20 H=His=Histidine: codons CAC, CAU

I=Ile=Isoleucine: codons AUA, AUC, AUU

K=Lys=Lysine: codons AAA, AAG

L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU

M=Met=Methionine: codon AUG

25 N=Asp=Asparagine: codons AAC, AAU

P=Pro=Proline: codons CCA, CCC, CCG, CCU

Q=Gln=Glutamine: codons CAA, CAG

R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU

S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU

30 T=Thr=Threonine: codons ACA, ACC, ACG, ACU

V=Val=Valine: codons GUA, GUC, GUG, GUU

W=Trp=Tryptophan: codon UGG

Y=Tyr=Tyrosine: codons UAC, UAU

Therefore, the present invention discloses codon redundancy which can

35 result in differing DNA molecules expressing an identical protein. For purposes of this specification, a sequence bearing one or more replaced

codons will be defined as a degenerate variation. Also included within the scope of this invention are mutations either in the DNA sequence or the translated protein which do not substantially alter the ultimate physical properties of the expressed protein. For example, substitution 5 of valine for leucine, arginine for lysine, or asparagine for glutamine may not cause a change in functionality of the polypeptide.

It is known that DNA sequences coding for a peptide can be altered so as to code for a peptide having properties that are different than those of the naturally occurring peptide. Methods of altering the 10 DNA sequences include but are not limited to site directed mutagenesis. Examples of altered properties include but are not limited to changes in the affinity of an enzyme for a substrate or a receptor for a ligand.

As used herein, a "biologically active equivalent" or "functional derivative" of a wild-type human HG38 possesses a biological 15 activity that is substantially similar to the biological activity of the wild type human HG38. The term "functional derivative" is intended to include the "fragments," "mutants," "variants," "degenerate variants," "analogs" and "homologues" or to "chemical derivatives" of the wild type human HG38 protein. The term "fragment" is meant to refer to any 20 polypeptide subset of wild-type human HG38. The term "mutant" is meant to refer to a molecule that may be substantially similar to the wild-type form but possesses distinguishing biological characteristics. Such altered characteristics include but are in no way limited to altered substrate binding, altered substrate affinity and altered sensitivity to 25 chemical compounds affecting biological activity of the human HG38 or human HG38 functional derivative. The term "variant" is meant to refer to a molecule substantially similar in structure and function to either the entire wild-type protein or to a fragment thereof. A molecule is "substantially similar" to a wild-type human HG38-like protein if both 30 molecules have substantially similar structures or if both molecules possess similar biological activity. Therefore, if the two molecules possess substantially similar activity, they are considered to be variants even if the structure of one of the molecules is not found in the other or even if the two amino acid sequences are not identical. The term 35 "analog" refers to a molecule substantially similar in function to either

the full-length human HG38 protein or to a biologically active fragment thereof.

As used herein in reference to a human HG38 gene or encoded protein, a "polymorphic" HG38 is an HG38 that is naturally found as an allele in the population at large. A polymorphic form of HG38 can have a different nucleotide sequence from the particular human HG38 allele disclosed herein. However, because of silent mutations, a polymorphic HG38 gene can encode the same or different amino acid sequence as that disclosed herein. Further, some polymorphic forms HG38 will exhibit biological characteristics that distinguish the form from wild-type receptor activity, in which case the polymorphic form is also a mutant.

A protein or fragment thereof is considered purified or isolated when it is obtained at a concentration at least about five-fold to ten-fold higher than that found in nature. A protein or fragment thereof is considered substantially pure if it is obtained at a concentration of at least about 100-fold higher than that found in nature. A protein or fragment thereof is considered essentially pure if it is obtained at a concentration of at least about 1000-fold higher than that found in nature.

Probes and Primers

The HG38 receptor disclosed herein shows a tissue specific pattern of expression. Therefore, polynucleotides of this invention can be used as probes for tissue typing. Polynucleotide probes comprising full length or partial sequences of SEQ ID NO:1 can be used to determine whether a tissue expresses HG38 RNA. The temporal and tissue specific expression of HG38 RNA throughout an animal can also be studied using polynucleotide probes. The effect of modulators that effect the transcription of the HG38 receptor gene can be studied via the use of these probes. A preferred probe is a single stranded antisense probe having at least the full length of the coding sequence of HG38. It is also preferred to use probes that have less than the full length sequence, and contain sequences highly specific for HG38 DNA or RNA.

A nucleotide probe is "highly specific" for HG38 DNA or RNA if one of skill in the art can use standard techniques to determine

hybridization and washing conditions through which one can detect an HG38 encoding DNA in a Southern Blot of total human genomic DNA (digested with a restriction enzyme to an average size of about 4000 nucleotides) without visually detectable nonspecific background

5 hybridization. A probe is specific if one can detect the HG38 DNA despite any visually detectable nonspecific background hybridization that may be present. The identification of a sequence(s) for use as a specific probe is well known in the art and involves choosing a sequence(s) that is unique to the target sequence, or is specific or highly specific thereto. It is
10 preferred that polynucleotides that are probes have at least about 25 nucleotides, more preferably about 30 to 35 nucleotides. The longer probes are believed to be more specific for HG38 genes and RNAs and can be used under more stringent hybridization conditions. Longer probes can be used but can be more difficult to prepare synthetically, or
15 can result in lower yields from a synthesis. Examples of sequences within SEQ ID NO:1 that are useful as probes or primers are the HG38 series of primers given in Example 1. However, one skilled in the art will recognize that these are only a few of the useful probe or primer sequences that can be derived from SEQ ID NO:1.

20 Polynucleotides having sequences that are unique or highly specific for HG38 can be used as primers in amplification reaction assays. These assays can be used in tissue typing as described herein. Additionally, amplification reactions employing primers derived from HG38 sequences can be used to obtain amplified HG38 DNA using the
25 HG38 DNA of the cells as an initial template. The HG38 DNA so obtained can be a mutant or polymorphic form of HG38 that differ from SEQ ID NO:1 by one or more nucleotides of the HG38 open reading frame or sequences flanking the ORF. The differences can be associated with a non-defective naturally occurring allele or with a defective form of HG38.
30 Thus, polynucleotides of this invention can be used in allelic identification of various HG38 genes or the detection of a defective HG38 gene.

35 Probes can be labeled by any number of ways known in the art including isotopes, enzymes, substrates, chemiluminescent, electrochemiluminescent, biotin and fret pairs among many others. A probe so labeled can generate a detectable signal directly (e.g., isotopes),

or upon hybridization (fret pairs), or indirectly after a chemical (e.g., luminescence) or biochemical reaction (e.g., enzyme-substrate) or after binding a strepavidin linked moiety that can generate a detectable signal directly or indirectly. The labeling of probes and the generation of 5 detectable signals are well known techniques in the art.

A primer is specific for the amplification of HG38 sequences if one of skill in the art can use standard techniques to determine conditions under which an amplification reaction yields a predominant amplified product corresponding to the HG38 sequences. A primer is 10 highly specific if no background amplification products are visually detectable.

Many types of amplification reactions are known in the art and include Polymerase Chain Reaction and Reverse Transcriptase Polymerase Chain Reaction (See e.g., PCR Primer, edited by 15 C.W.Dieffenbach and G.S.Dveksler, (1995). Cold Spring Harbor Laboratory Press.), Strand Displacement Amplification, Self-Sustained Sequence Reaction, and any other amplification known to one of skill in the art that uses primers. Any of these or like reactions can be used with primers derived from SEQ ID NO:1.

20

Polynucleotide Cloning

The HG38 nucleotide and amino acid sequences provided herein can be used to isolate and/or clone HG38 polynucleotides. Any of a variety of procedures can be used to clone HG38. These methods 25 include, but are not limited to, (1) a RACE PCR cloning technique (Frohman, et al., 1988, *Proc. Natl. Acad. Sci.* 85: 8998-9002). 5' and/or 3' RACE can be performed to generate a full length cDNA sequence. This strategy involves using gene-specific oligonucleotide primers for PCR amplification of HG38 cDNA. These gene-specific primers are designed 30 through identification of an expressed sequence tag (EST) nucleotide sequence which has been identified by searching any number of publicly available nucleic acid and protein databases; (2) direct functional expression of the HG38 cDNA following the construction of an HG38-containing cDNA library in an appropriate expression vector system; (3) 35 screening a HG38-containing cDNA library constructed in a bacteriophage or plasmid shuttle vector with a labeled degenerate

oligonucleotide probe designed from the amino acid sequence of the HG38 protein; (4) screening a HG38-containing cDNA library constructed in a bacteriophage or plasmid shuttle vector with a partial cDNA encoding the HG38 protein. This partial cDNA is obtained by the

5 specific PCR amplification of HG38 DNA fragments through the design of degenerate oligonucleotide primers from the amino acid sequence known for other receptors which are related to the HG38 protein (e.g., leuteneizing, follicle-stimulating and thyroid stimulating hormone receptors); (5) screening an HG38-containing cDNA library constructed

10 in a bacteriophage or plasmid shuttle vector with a partial cDNA encoding the HG38 protein. This strategy can also involve using gene-specific oligonucleotide primers for PCR amplification of HG38 cDNA identified as an EST as described herein; or (6) designing 5' and 3' gene specific oligonucleotides using SEQ ID NO:1 as a template so that either

15 the full length cDNA can be generated by known PCR techniques, or a portion of the coding region can be generated by these same known PCR techniques to generate and isolate a portion of the coding region to use as a probe to screen one of numerous types of cDNA and/or genomic libraries in order to isolate a full length version of the nucleotide

20 sequence encoding HG38.

It is readily apparent to those skilled in the art that other types of libraries, as well as libraries constructed from other cells types or species types, can be useful for isolating a human HG38-encoding DNA, a mammalian HG38 homologue, or mutant or polymorphic forms of HG38 receptor DNA or RNA. Other types of libraries include, but are not limited to, cDNA libraries derived from other cells or cell lines other than human cells or tissue such as primate, murine, rodent, porcine and bovine cells or any other such vertebrate host which contains HG38-encoding DNA. Additionally, an HG38 gene can be isolated by

25 oligonucleotide- or polynucleotide- based hybridization screening of a vertebrate genomic library, including but not limited to primate, murine, rodent, porcine or bovine genomic libraries, as well as concomitant human genomic DNA libraries.

30

It is readily apparent to those skilled in the art that suitable

35 cDNA libraries can be prepared from cells or cell lines which express an HG38 receptor. The selection of cells or cell lines for use in preparing a

cDNA library to isolate a HG38 cDNA can be done by first detecting cell associated HG38 receptors using an assay for HG38, *e.g.*, an assay using antibodies disclosed herein or a PCR assay using HG38-specific primers.

Preparation of cDNA libraries can be performed by

5 standard techniques well known in the art. Well known cDNA library construction techniques can be found for example, in Sambrook, et al., 1989, *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. Complementary DNA libraries can also be obtained from numerous commercial sources,

10 including but not limited to Clontech Laboratories, Inc., Palo Alto, CA, USA and Stratagene, Inc., La Jolla, CA, USA.

It is also readily apparent to those skilled in the art that DNA encoding HG38 can also be isolated from a suitable genomic DNA library. Construction of genomic DNA libraries can be performed by 15 standard techniques well known in the art. Well known genomic DNA library construction techniques can be found in Sambrook, et al., *supra*.

In order to clone the HG38 gene by one of the preferred methods, the amino acid sequence or DNA sequence of HG38 or a homologous protein may be necessary. To accomplish this, the HG38 or 20 a homologous protein can be purified, *e.g.*, through cross reaction with the anti-HG38 antibodies taught herein, and partial amino acid sequence(s) determined by automated sequenators. It is not necessary to determine the entire amino acid sequence, but the linear sequence of two regions of 6 to 8 amino acids can be determined for the PCR 25 amplification of a partial HG38 DNA fragment. Once suitable amino acid sequences have been identified, the DNA sequences capable of encoding them are synthesized. Because the genetic code is degenerate, more than one codon can be used to encode a particular amino acid, and therefore, the amino acid sequence can be encoded by any of a set of 30 similar, degenerate, DNA oligonucleotides. Only one member of the degenerate set will be identical to the HG38 sequence but others in the set will be capable of hybridizing to HG38 DNA even in the presence of DNA oligonucleotides with mismatches. The mismatched DNA oligonucleotides can still sufficiently hybridize to the HG38 DNA to 35 permit identification and isolation of HG38 encoding DNA. Alternatively, the nucleotide sequence of a region of an expressed

sequence can be identified by searching one or more available genomic databases. Gene-specific primers can be used to perform PCR amplification of a cDNA of interest from either a cDNA library or a population of cDNAs. As noted herein, the appropriate nucleotide 5 sequence for use in a PCR-based method can be obtained from SEQ ID NO:1, either for the purpose of isolating overlapping 5' and 3' PCR products for generation of a full-length sequence coding for HG38, or to isolate a portion of the nucleotide sequence coding for HG38 for use as a probe to screen one or more cDNA- or genomic-based libraries to isolate 10 a full-length sequence encoding HG38 or HG38-like proteins.

In a method used in Example 1, the HG38 full length cDNA of the present invention was generated by a method of cDNA screening called Reduced Complexity cDNA Analysis (RCCA). Briefly, the extension of partial cDNA sequences have historically been achieved 15 with one or both of the two commonly used methods: filter screening of cDNA libraries by hybridization with labeled probes, and 5'- and 3'-RACE with total cellular mRNA by PCR. The first method is effective but laborious and slow while the latter method is fast but limited in efficiency. This RACE protocol is hindered by limited length of extension 20 due to the use of the entire cellular mRNA population in a single reaction. Since smaller fragments are amplified much more efficiently than larger fragments by PCR in the same reaction, PCR products obtained using the second method are often quite small.

The RCCA method improves upon known methods of cDNA 25 library screening by initially constructing and subdividing cDNA libraries followed by isolating 5'- and 3'- flanking fragments by PCR. Since each pool is unlikely to contain more than one clone for a given gene which is low to moderately expressed, competition between large and small PCR products in one pool does not exist, making it possible to 30 isolate fragments of various sizes. One definite advantage of the method as described herein is the efficiency, throughput, and its potential to isolate alternatively spliced cDNA forms.

The RCCA process provides for rapid extension of a partial cDNA sequence based on subdividing a primary cDNA library and DNA 35 amplification by polymerase chain reaction (PCR). A cDNA library is constructed with cDNA primed by random, oligo-dT or a combination of

both random and oligo-dT primers and then subdivided into pools at approximately 10,000 -20,000 clones per pool ("superpools"). Each superpool is amplified separately and therefore represents an independent portion of the cDNA molecules from the original mRNA

5 source. Samples from all the superpools are collected and transferred into 96-well plates. To extend a partial cDNA sequence, such as SEQ ID NO:1, positive pools containing the partial cDNA sequence are first identified by PCR with a pair of primers complementary to the cDNA sequence. Each positive pool in the library contains an independent

10 clone of the cDNA sequence; within each clone are embedded the partial cDNA sequence and its flanking fragments. The flanking fragments are isolated by PCR with primers complementary to the known vector and cDNA sequences and then sequenced directly. The DNA sequences from these fragments plus the original partial cDNA sequence are

15 assembled into a continuous fragment, resulting in the extension of the partial cDNA sequence and the eventual determination of its full-length gene sequence by repeating the process, if necessary, until a complete open reading frame is obtained.

20 The fundamental principle of this process is to subdivide a complex library into superpools of about 10,000 to about 20,000 clones. A library of two million primary clones, a number large enough to cover most mRNA transcripts expressed in the tissue, can be subdivided into 188 pools and stored in two 96-well plates. Since the number of transcripts for most genes is fewer than one copy per ~10,000 transcripts

25 in total cellular mRNA, each pool is unlikely to contain more than one clone for a given cDNA sequence. Such reduced complexity makes it possible to use PCR to isolate flanking fragments of partial cDNA sequences larger than those obtained by known methods.

30 The skilled artisan, aided with this specification, will understand the far reaching cDNA cloning process disclosed herein: multiple primer combinations from an EST or other partial cDNA sequence, in combination with flanking vector primer oligonucleotides can be used to "walk" in both directions away from the internal, gene specific, sequence, and respective primers, such that a contig

35 representing a full length cDNA can be constructed. This procedure relies on the ability to screen multiple pools which comprise a

representative portion of the total cDNA library. This procedure is not dependent upon using a cDNA library with directionally cloned inserts. Instead, both 5' and 3' vector and gene specific primers are added and a contig map is constructed from additional screening of positive pools

5 using both vector primers and gene specific primers. Of course, these gene specific primers are initially constructed from a known nucleic acid fragment such as an expressed sequence tag. However, as the walk continues, gene specific primers are utilized from the 5' and 3' boundaries of the newly identified regions of the cDNA. As the walk

10 continues, there is still no requirement that the vector orientation of a yet unidentified fragment be known. Instead, all combinations are tested on a positive pool and the actual vector orientation is determined by the ability of certain vector/gene specific primers to generate the predicted PCR fragment. A full-length cDNA can then be easily constructed by

15 known subcloning procedures.

Isolation of other species homologs of the HG38 gene

The HG38 gene from different species, *e.g.* mouse, rat, dog, are isolated by screening of a cDNA library with portions of the gene that

20 have been obtained from cDNA of the species of interest using PCR primers designed from the human HG38 sequence. Degenerate PCR is performed by designing primers of 17-20 nucleotides with 32-128 fold degeneracy by selecting regions that code for amino acids that have low codon degeneracy *e.g.* Met and Trp. When selecting these primers

25 preference is given to regions that are conserved in the protein. PCR products are analyzed by DNA sequence analysis to confirm their similarity to human HG38. The correct product is used to screen cDNA libraries by colony or plaque hybridization at high stringency.

Alternatively, probes derived directly from the human HG38 gene are

30 utilized to isolate the cDNA sequence of HG38 from different species by hybridization at reduced stringency. A cDNA library can be generated as known in the art or as described herein.

Transgenic Animals

In reference to the transgenic animals of this invention, we refer to transgenes and genes. As used herein, a "transgene" is a genetic construct including a gene. The transgene is integrated into one or more chromosomes in the cells in an animal or its ancestor by methods known in the art. Once integrated, the transgene is carried in at least one place in the chromosomes of a transgenic animal. A gene is a nucleotide sequence that encodes a protein. The gene and/or transgene can also include genetic regulatory elements and/or structural elements known in the art.

The term "animal" is used herein to include all mammals, except humans. It also includes an individual animal in all stages of development, including embryonic and fetal stages. Preferably the animal is a rodent, and most preferably mouse or rat. A "transgenic animal" is an animal containing one or more cells bearing genetic information received, directly or indirectly, by deliberate genetic manipulation at a subcellular level, such as by microinjection or infection with recombinant virus. This introduced DNA molecule can be integrated within a chromosome, or it can be extra-chromosomally replicating DNA. Unless otherwise noted or understood from the context of the description of an animal, the term "transgenic animal" as used herein refers to a transgenic animal in which the genetic information was introduced into a germ line cell, thereby conferring the ability to transfer the information to offspring. If offspring in fact possess some or all of the genetic information, then they, too, are transgenic animals. The genetic information is typically provided in the form of a transgene carried by the transgenic animal.

The genetic information received by the non-human animal can be foreign to the species of animal to which the recipient belongs, or foreign only to the particular individual recipient. In the last case, the information can be altered or it can be expressed differently than the native gene. Alternatively, the altered or introduced gene can cause the native gene to become non-functional to produce a "knockout" animal.

As used herein, a "targeted gene" or "Knockout" (KO) transgene is a DNA sequence introduced into the germline of a non-human animal by way of human intervention, including but not limited



to, the methods described herein. The targeted genes of the invention include nucleic acid sequences which are designed to specifically alter cognate endogenous alleles of the non-human animal.

An altered HG38 receptor gene should not fully encode the 5 same receptor endogenous to the host animal, and its expression product can be altered to a minor or great degree, or absent altogether. In cases where it is useful to express a non-native HG38 receptor in a transgenic animal in the absence of a endogenous HG38 receptor we prefer that the altered HG38 gene induce a null, "knockout," phenotype 10 in the animal. However a more modestly modified HG38 gene can also be useful and is within the scope of the present invention.

A type of target cell for transgene introduction is the embryonic stem cell (ES). ES cells can be obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (M. J. 15 Evans *et al.*, *Nature* 292:154-156 (1981); Bradley *et al.*, *Nature* 309:255-258 (1984); Gossler *et al.* *Proc. Natl. Acad. Sci. USA* 83:9065-9069 (1986); and Robertson *et al.*, *Nature* 322:445-448 (1986)). Transgenes can be 20 efficiently introduced into the ES cells by a variety of standard techniques such as DNA transfection, microinjection, or by retrovirus-mediated transduction. The resultant transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The introduced ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal (R. Jaenisch, *Science* 240: 1468-1474 (1988)). Animals are screened for those resulting in germline 25 transforms. These are crossed to produce animals homozygous for the transgene.

Methods for evaluating the targeted recombination events as well as the resulting knockout mice are readily available and known in the art. Such methods include, but are not limited to DNA (Southern) 30 hybridization to detect the targeted allele, polymerase chain reaction (PCR), polyacrylamide gel electrophoresis (PAGE) and Western blots to detect DNA, RNA and protein.

This may have a therapeutic aim. (Gene therapy is 35 discussed below.) The presence of a mutant, allele or variant sequence within cells of an organism, particularly when in place of a homologous endogenous sequence, may allow the organism to be used as a model in

testing and/or studying the role of the HG38 gene or substances which modulate activity of the encoded polypeptide and/or promoter *in vitro* or are otherwise indicated to be of therapeutic potential.

5 Expression of HG38

The present invention also relates to recombinant vectors and recombinant hosts, both prokaryotic and eukaryotic, which contain the substantially purified nucleic acid molecules disclosed throughout this specification.

10 Therefore, the present invention also relates to methods of expressing HG38 and biological equivalents disclosed herein, assays employing these recombinantly expressed gene products, cells expressing these gene products, and modulators, agonistic and/or antagonistic compounds identified through the use of assays utilizing
15 these recombinant forms, including, but not limited to, one or more compounds or molecules that act through direct contact with the receptor, particularly with the ligand binding domain, or through direct or indirect contact with a ligand which either interacts with the receptor or with the transcription or translation of HG38, thereby modulating
20 HG38 expression.

A variety of expression vectors can be used to express recombinant HG38 in host cells. Expression vectors are defined herein as DNA sequences that are required for the transcription of cloned DNA and the translation of their mRNAs in an appropriate host. Such
25 vectors can be used to express eukaryotic DNA in a variety of hosts such as bacteria, bluegreen algae, plant cells, insect cells and animal cells. Specifically designed vectors allow the shuttling of DNA between hosts such as bacteria-yeast or bacteria-animal cells. An appropriately constructed expression vector should contain: an origin of replication for
30 autonomous replication in host cells, selectable markers, a limited number of useful restriction enzyme sites, a potential for high copy number, and active promoters. A promoter is defined as a DNA sequence that directs RNA polymerase to bind to DNA and initiate RNA synthesis. A strong promoter is one which causes mRNAs to be
35 initiated at high frequency. Expression vectors can include, but are not

limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses.

Commercially available mammalian expression vectors which can be suitable for recombinant human HG38 expression, include 5 but are not limited to, pcDNA3.1 (Invitrogen), pLITMUS28, pLITMUS29, pLITMUS38 and pLITMUS39 (New England Biolabs), pcDNAI, pcDNAIamp (Invitrogen), pcDNA3 (Invitrogen), pMC1neo (Stratagene), pXT1 (Stratagene), pSG5 (Stratagene), EBO-pSV2-neo (ATCC 37593) pBPV-1(8-2) (ATCC 37110), pdBPV-MMTneo(342-12) (ATCC 37224), 10 pRSVgpt (ATCC 37199), pRSVneo (ATCC 37198), pSV2-dhfr (ATCC 37146), pUCTag (ATCC 37460), and lZD35 (ATCC 37565).

A variety of bacterial expression vectors can be used to express recombinant human HG38 in bacterial cells. Commercially available bacterial expression vectors which are suitable for 15 recombinant human HG38 expression include, but are not limited to pQE (Qiagen), pET11a (Novagen), lambda gt11 (Invitrogen), and pKK223-3 (Pharmacia).

A variety of fungal cell expression vectors can be used to express recombinant human HG38 in fungal cells. Commercially 20 available fungal cell expression vectors which are suitable for recombinant human HG38 expression include but are not limited to pYES2 (Invitrogen) and *Pichia* expression vector (Invitrogen).

A variety of insect cell expression vectors can be used to express recombinant receptor in insect cells. Commercially available 25 insect cell expression vectors which are suitable for recombinant expression of human HG38 include but are not limited to pBlueBacIII and pBlueBacHis2 (Invitrogen), and pAcG2T (Pharmingen).

An expression vector containing DNA encoding a human HG38-like protein can be used for expression of human HG38 in a 30 recombinant host cell. Recombinant host cells can be prokaryotic or eukaryotic, including but not limited to bacteria such as *E. coli*, fungal cells such as yeast, mammalian cells including but not limited to cell lines of human, bovine, porcine, monkey and rodent origin, and insect cells including but not limited to *Drosophila*- and silkworm-derived cell 35 lines. Cell lines derived from mammalian species which can be suitable and which are commercially available, include but are not limited to, L

cells L-M(TK⁻) (ATCC CCL 1.3), L cells L-M (ATCC CCL 1.2), Saos-2 (ATCC HTB-85), 293 (ATCC CRL 1573), Raji (ATCC CCL 86), CV-1 (ATCC CCL 70), COS-1 (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C127I (ATCC CRL 1616), BS-C-1 (ATCC CCL 26), MRC-5 (ATCC CCL 171) and CPAE (ATCC CCL 209).

5 The expression vector can be introduced into host cells via any one of a number of techniques including but not limited to transformation, transfection, protoplast fusion, and electroporation.

10 The expression vector-containing cells are individually analyzed to determine whether they produce human HG38 protein. Identification of human HG38 expressing cells can be done by several means, including but not limited to immunological reactivity with anti-human HG38 antibodies, labeled ligand binding and the presence of host cell-15 associated human HG38 activity.

The cloned human HG38 cDNA obtained through the methods described herein can be recombinantly expressed by molecular cloning into an expression vector (such as pcDNA3.1, pQE, pBlueBacHis2 and pLITMUS28) containing a suitable promoter and 20 other appropriate transcription regulatory elements, and transferred into prokaryotic or eukaryotic host cells to produce recombinant human HG38. Techniques for such manipulations can be found described in Sambrook, et al., *supra*, and are well known and easily available to the one of ordinary skill in the art.

25 Expression of human HG38 DNA can also be performed using *in vitro* produced synthetic mRNA. Synthetic mRNA can be efficiently translated in various cell-free systems, including but not limited to wheat germ extracts and reticulocyte extracts, as well as efficiently translated in cell based systems, including but not limited to 30 microinjection into frog oocytes, with microinjection into frog oocytes being preferred.

To determine the human HG38 cDNA sequence(s) that yields optimal levels of human HG38, cDNA molecules including but not limited to the following can be constructed: a cDNA fragment containing 35 the full-length open reading frame for human HG38 as well as various constructs containing portions of the cDNA encoding only specific

domains of the protein or rearranged domains of the protein. All constructs can be designed to contain none, all or portions of the 5' and/or 3' untranslated region of a human HG38 cDNA. The expression levels and activity of human HG38 can be determined following the 5 introduction, both singly and in combination, of these constructs into appropriate host cells. Following determination of the human HG38 cDNA cassette yielding optimal expression in transient assays, this HG38 cDNA construct is transferred to a variety of expression vectors (including recombinant viruses), including but not limited to those for 10 mammalian cells, plant cells, insect cells, oocytes, bacteria, and yeast cells.

Following expression of HG38 in a host cell, HG38 polypeptides can be recovered. Several HG38 protein purification procedures are available and suitable for use. HG38 protein and 15 polypeptides can be purified from cell lysates and extracts, or from conditioned culture medium, by various combinations of, or individual application of methods including ultrafiltration, acid extraction, alcohol precipitation, salt fractionation, ionic exchange chromatography, phosphocellulose chromatography, lecithin chromatography, affinity 20 (e.g., antibody or His-Ni) chromatography, size exclusion chromatography, hydroxylapatite adsorption chromatography and chromatography based on hydrophobic or hydrophilic interactions. In some instances, protein denaturation and refolding steps can be employed. High performance liquid chromatography (HPLC) and 25 reversed phase HPLC can also be useful. Dialysis can be used to adjust the final buffer composition.

Anti-HG38 Antibodies

The present invention also relates to polyclonal and 30 monoclonal antibodies raised in response to either the human form of HG38 disclosed herein, or a biologically active fragment thereof. It will be especially preferable to raise antibodies against epitopes within the NH₂-terminal domain or the extracellular inter-membrane domains of HG38. It is also preferable to raise antibodies to epitopes which show the 35 least homology to other known glycoprotein hormone receptor proteins.

An antibody is specific for an HG38 epitope if one of skill in the art can use standard techniques to determine conditions under which one can detect HG38 in a Western Blot of a sample from a host cell that displays HG38 on its surface. The blot can be of a native or

5 denaturing gel as appropriate for the epitope. An antibody is highly specific for an HG38 epitope if no nonspecific background binding is visually detectable. An antibody can also be considered highly specific for HG38 if the binding of the antibody to HG38 can not be competed by non-HG38 peptides, polypeptides or proteins.

10 Recombinant HG38 protein can be separated from other cellular proteins by use of an immunoaffinity column made with monoclonal or polyclonal antibodies specific for full-length HG38 protein, or polypeptide fragments of HG38 protein. Additionally, polyclonal or monoclonal antibodies can be raised against a synthetic 15 peptide (usually from about 9 to about 25 amino acids in length) from a portion of the protein as disclosed in SEQ ID NO:2. Monospecific antibodies to human HG38 are purified from mammalian antisera containing antibodies reactive against human HG38 or are prepared as monoclonal antibodies reactive with human HG38 using the technique of 20 Kohler and Milstein (1975, *Nature* 256: 495-497). Monospecific antibody as used herein is defined as a single antibody species or multiple antibody species with homogenous binding characteristics for human HG38. Homogenous binding as used herein refers to the ability of the antibody species to bind to a specific antigen or epitope, such as those 25 associated with human HG38, as described herein. Human HG38-specific antibodies are raised by immunizing animals such as mice, rats, guinea pigs, rabbits, goats, horses and the like, with an appropriate concentration of human HG38 protein or a synthetic peptide generated from a portion of human HG38 with or without an immune 30 adjuvant.

Preimmune serum is collected prior to the first immunization. Each animal receives between about 0.1 mg and about 1000 mg of human HG38 protein associated with an acceptable immune adjuvant. Such acceptable adjuvants include, but are not limited to, 35 Freund's complete, Freund's incomplete, alum-precipitate, water in oil emulsion containing *Corynebacterium parvum* and tRNA. The initial

immunization consists of human HG38 protein or peptide fragment thereof in, preferably, Freund's complete adjuvant at multiple sites either subcutaneously (SC), intraperitoneally (IP) or both. Each animal is bled at regular intervals, preferably weekly, to determine antibody 5 titer. The animals may or may not receive booster injections following the initial immunization. Those animals receiving booster injections are generally given an equal amount of human HG38 in Freund's incomplete adjuvant by the same route. Booster injections are given at about three week intervals until maximal titers are obtained. At about 7 10 days after each booster immunization or about weekly after a single immunization, the animals are bled, the serum collected, and aliquots are stored at about -20°C.

Monoclonal antibodies (mAb) reactive with human HG38 are prepared by immunizing inbred mice, preferably Balb/c, with 15 human HG38 protein. The mice are immunized by the IP or SC route with about 1 mg to about 100 mg, preferably about 10 mg, of human HG38 protein in about 0.5 ml buffer or saline incorporated in an equal volume of an acceptable adjuvant, as discussed herein. Freund's complete adjuvant is preferred. The mice receive an initial immunization on day 20 0 and are rested for about 3 to about 30 weeks. Immunized mice are given one or more booster immunizations of about 1 to about 100 mg of human HG38 in a buffer solution such as phosphate buffered saline by the intravenous (IV) route. Lymphocytes, from antibody positive mice, preferably splenic lymphocytes, are obtained by removing spleens from 25 immunized mice by standard procedures known in the art. Hybridoma cells are produced by mixing the splenic lymphocytes with an appropriate fusion partner, preferably myeloma cells, under conditions which will allow the formation of stable hybridomas. Fusion partners can include, but are not limited to: mouse myelomas P3/NS1/Ag 4-1; 30 MPC-11; S-194 and Sp 2/0, with Sp 2/0 being preferred. The antibody producing cells and myeloma cells are fused in polyethylene glycol, about 1000 mol. wt., at concentrations from about 30% to about 50%. Fused hybridoma cells are selected by growth in hypoxanthine, 35 thymidine and aminopterin supplemented Dulbecco's Modified Eagles Medium (DMEM) by procedures known in the art. Supernatant fluids are collected from growth positive wells on about days 14, 18, and 21 and

are screened for antibody production by an immunoassay such as solid phase immunoradioassay (SPIRA) using human HG38 as the antigen. The culture fluids are also tested in the Ouchterlony precipitation assay to determine the isotype of the mAb. Hybridoma cells from antibody 5 positive wells are cloned by a technique such as the soft agar technique of MacPherson, 1973, *Soft Agar Techniques*, in *Tissue Culture Methods and Applications*, Kruse and Paterson, Eds., Academic Press.

Monoclonal antibodies are produced *in vivo* by injection of pristine primed Balb/c mice, approximately 0.5 ml per mouse, with 10 about 2×10^6 to about 6×10^6 hybridoma cells about 4 days after priming. Ascites fluid is collected at approximately 8-12 days after cell transfer and the monoclonal antibodies are purified by techniques known in the art.

In vitro production of anti-human HG38 mAb is carried out 15 by growing the hybridoma in DMEM containing about 2% fetal calf serum to obtain sufficient quantities of the specific mAb. The mAb are purified by techniques known in the art.

Antibody titers of ascites or hybridoma culture fluids are determined by various serological or immunological assays which 20 include, but are not limited to, precipitation, passive agglutination, enzyme-linked immunosorbent antibody (ELISA) technique and radioimmunoassay (RIA) techniques. Similar assays are used to detect the presence of human HG38 in body fluids or tissue and cell extracts.

It is readily apparent to those skilled in the art that the 25 herein described methods for producing monospecific antibodies can be utilized to produce antibodies specific for human HG38 peptide fragments, or full-length human HG38.

Human HG38 antibody affinity columns are made, for example, by adding the antibodies to Affigel-10 (Biorad), a gel support 30 which is pre-activated with N-hydroxysuccinimide esters such that the antibodies form covalent linkages with the agarose gel bead support. The antibodies are then coupled to the gel via amide bonds with the spacer arm. The remaining activated esters are then quenched with 1M ethanolamine HCl (pH 8). The column is washed with water followed by 35 0.23 M glycine HCl (pH 2.6) to remove any non-conjugated antibody or extraneous protein. The column is then equilibrated in phosphate

buffered saline (pH 7.3) and the cell culture supernatants or cell extracts containing full-length human HG38 or human HG38 protein fragments are slowly passed through the column. The column is then washed with phosphate buffered saline until the optical density (A_{280}) falls to

5 background, then the protein is eluted with 0.23 M glycine-HCl (pH 2.6). The purified human HG38 protein is then dialyzed against phosphate buffered saline.

Levels of human HG38 in host cells is quantified by a variety of techniques including, but not limited to, immunoaffinity and/or 10 ligand affinity techniques. HG38-specific affinity beads or HG38-specific antibodies are used to isolate 35 S-methionine labeled or unlabelled HG38. Labeled HG38 protein is analyzed by SDS-PAGE. Unlabelled HG38 protein is detected by Western blotting, ELISA or RIA assays employing either HG38 protein specific antibodies and/or antiphosphotyrosine 15 antibodies.

Modulators, Agonists and Antagonists of HG38

The present invention is also directed to methods for screening for compounds or molecules which modulate the expression 20 of DNA or RNA encoding a human HG38 protein. Compounds or molecules which modulate these activities can be DNA, RNA, peptides, proteins, or non-proteinaceous organic molecules. They can modulate by increasing or attenuating the expression of DNA or RNA encoding human HG38. Compounds that modulate the expression of DNA or 25 RNA encoding human HG38 or are agonists or antagonists of the biological function thereof can be detected by a variety of assays. The assay can be a simple "yes/no" assay to determine whether there is a change in expression or function. The assay can be made quantitative by comparing the expression or function of a test sample with the levels of 30 expression or function in a standard sample. Kits containing human HG38, antibodies to human HG38, or modified human HG38 can be prepared by known methods for such uses.

The DNA molecules, RNA molecules, recombinant protein and antibodies of the present invention can be used to screen and 35 measure levels of human HG38. The recombinant proteins, DNA molecules, RNA molecules and antibodies lend themselves to the

formulation of kits suitable for the detection and typing of human HG38. Such a kit would comprise a compartmentalized carrier suitable to hold in close confinement at least one container. The carrier would further comprise reagents such as recombinant HG38 or anti-HG38 antibodies 5 suitable for detecting human HG38. The carrier can also contain a means for detection such as labeled antigen or enzyme substrates or the like.

Pharmaceutical Compositions

10 Pharmaceutically useful compositions comprising agonists, antagonist or modulators of human HG38 can be formulated according to known methods such as by the admixture of a pharmaceutically acceptable carrier. Examples of such carriers and methods of formulation can be found in Remington's Pharmaceutical 15 Sciences. To form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of the protein, DNA, RNA, modified human HG38, or either HG38 modulators, agonsits or antagonists.

20 Therapeutic or diagnostic compositions of the invention are administered to an individual in amounts sufficient to treat or diagnose disorders. The effective amount can vary according to a variety of factors such as the individual's condition, weight, sex and age. Other factors include the mode of administration.

25 The pharmaceutical compositions can be provided to the individual by a variety of routes such as subcutaneous, topical, oral and intramuscular.

30 The term "chemical derivative" describes a molecule that contains additional chemical moieties which are not normally a part of the base molecule. Such moieties can improve the solubility, half-life, absorption, etc. of the base molecule. Alternatively the moieties can attenuate undesirable side effects of the base molecule or decrease the toxicity of the base molecule. Examples of such moieties are described in a variety of texts, such as Remington's Pharmaceutical Sciences.

35 Compounds identified according to the methods disclosed herein can be used alone at appropriate dosages. Alternatively, co-

administration or sequential administration of other agents can be desirable.

The present invention also provides a means to obtain suitable topical, oral, systemic and parenteral pharmaceutical formulations for use in the methods of treatment of the present invention. The compositions containing compounds or molecules identified according to this invention as the active ingredient can be administered in a wide variety of therapeutic dosage forms in conventional vehicles for administration. For example, the compounds can be administered in such oral dosage forms as tablets, capsules (each including timed release and sustained release formulations), pills, powders, granules, elixirs, tinctures, solutions, suspensions, syrups and emulsions, or by injection. Likewise, they can also be administered in intravenous (both bolus and infusion), intraperitoneal, subcutaneous, topical with or without occlusion, or intramuscular form, all using forms well known to those of ordinary skill in the pharmaceutical arts.

Advantageously, compounds of the present invention can be administered in a single daily dose, or the total daily dosage can be administered in divided doses of two, three or four times daily. Furthermore, compounds for the present invention can be administered in intranasal form via topical use of suitable intranasal vehicles, or via transdermal routes, using those forms of transdermal skin patches well known to those of ordinary skill in that art. To be administered in the form of a transdermal delivery system, the dosage administration will, of course, be continuous rather than intermittent throughout the dosage regimen.

For combination treatment with more than one active agent, where the active agents are in separate dosage formulations, the active agents can be administered concurrently, or they each can be administered at separately staggered times.

The dosage regimen utilizing the compounds of the present invention is selected in accordance with a variety of factors including type, species, age, weight, sex and medical condition of the patient; the severity of the condition to be treated; the route of administration; the renal, hepatic and cardiovascular function of the patient; and the particular compound thereof employed. A physician or veterinarian of

ordinary skill can readily determine and prescribe the effective amount of the drug required to prevent, counter or arrest the progress of the condition. Optimal precision in achieving concentrations of drug within the range that yields efficacy without toxicity requires a regimen based
5 on the kinetics of the drug's availability to target sites. This involves a consideration of the distribution, equilibrium, and elimination of a drug.

The following examples are presented by the way of illustration and, because various other embodiments will be apparent to those in the art, the following is not to be construed as a limitation on the
10 scope of the invention:

EXAMPLE 1

Isolation of the HG38 receptor cDNA

Identification of a partial cDNA for the HG38 receptor

15 Polypeptide sequences of human G-protein coupled glycoprotein hormone receptors were used as probes to search the EST database dbEST of NCBI (National Center for Biotechnology Information) using the search program tFASTA. The sequences chosen were the protein sequences of known human receptors, i.e., receptors for
20 FSH (Follicle-stimulating hormone), TSH (thyroid-stimulating hormone), LH (leutinizing hormone). An EST (accession #aa424098) was found to be able to encode a polypeptide that is approximately 30% identical to these receptors at the amino acid level. This EST, containing a sequence of 493 base pairs, was sequenced from the 5' end of a clone
25 from a total human fetus cDNA library (the I.M.A.G.E. ID of this clone = 759936).

30 The DNA sequence information of this EST was used to isolate cDNA fragments containing the original EST. DNA sequences of these fragments were then determined and analyzed, resulting in the identification of the full-length coding sequence of the HG38 gene. The full-length cDNA sequence was then cloned into a bacterial vector.

Primers

The following primers were used for the isolation of HG38 as described below. For convenience and clarity, the SEQ ID NOS are presented here. In the following description, primers can be referred to 5 by the numerical component of their designation.

	HG38.179F	TGACGACATGGTGCCTGGTG	(SEQ ID NO:3)
	HG38.439R	GGCAAAGGCAGGCAGAGAGG	(SEQ ID NO:4)
	HG38.37F	CAACATCAGTCAGCTGCTCCG	(SEQ ID NO:5)
10	HG38.111R	CCCTTGGGAATGTATGTCAGAG	(SEQ ID NO:6)
	HG38.755F	ACAGCACTGGTAAGCATAAGGC	(SEQ ID NO:7)
	HG38.144F	CTGGCTGGTGTGGATGCGTT	(SEQ ID NO:8)
	HG38.3419	CCCATGGATCACAGCCTCTACC	(SEQ ID NO:9)
	HG38.99FS	CGCCGTGGGTCAGGAAC	(SEQ ID NO:10)
15	PBS.543R	GGGGATGTGCTGCAAGGCGA	(SEQ ID NO:11)
	PBS.873F	CCCAGGCTTACACTTATGCTTCC	(SEQ ID NO:12)

Cloning and sequencing of HG38

The full-length sequence of HG38 was isolated from a 20 placental cDNA library by multiple rounds RCCA (Reduced Complexity cDNA Analysis, described herein). Random and oligo dT primed cDNA librares of fetal brain, placenta, testes, and prostate consisting of approximately 4 million primary clones each was constructed in the plasmid vector pBluescript SK- (Stratagene, La Jolla, CA). The primary 25 clones of each library were subdivided into 188 superpools with each pool containing ~20,000 clones. Each pool was amplified separately and the resulting plasmid pools were collected and transferred into eight 96-well plates. Each 96-well plate was pooled into 8 "superpools," generating a total of 64 superpools covering the four libraries.

30 For the initial superpool scanning of the cDNA library, 5' and 3' primers predicted to be specific for the HG38 EST aa424098, (primers 179F +439R), as well as oligonucleotide primers both 5' and 3' of the polylinker sequence of the vector (primers PBS.873F and PBS.543R) were used. PCR reactions were carried out with AmpliTaq Gold (Perkin 35 Elmer-Roche, Branchberg, NJ, U.S.A) using standard PCR conditions as suggested by the enzyme supplier. The positive superpools were

scanned again by PCR with both pBluescript primers as well as the 5' and 3' primers associated with the HG38 EST.

Following superpool scanning, tissue specific scanning was done on the placental cDNA plasmid library using the same set of 5 primers. After positive wells were identified, insert-vector PCR (polymerase chain reaction) was carried out on the vectors from the positive pools using the following primer combinations: 179F+PBS.873F; 439R+PBS.543R. The fragments synthesized via the PCR were sequenced and assembled into a contiguous sequence.

10 Based on the new sequences, 3' and 5' HG38 primers 755F and 144F were synthesized for extension reactions. The following insert-vector primer combinations were used: 755F+PBS.543R, 755F+PBS.873F; 144R+PBS.543R, 144R+PBS.873F in the PCR reactions. PCR products were then sequenced and assembled with the previous sequence into a 15 contiguous sequence. Based on the new sequence two new primers 37F and 111R were designed and synthesized.

20 To identify new pools for the 5' end of the gene these new primers were used to scan the placental cDNA plasmid library. After identification of positive pools, insert-vector PCR was carried out using the following primer combination: 111R+PBS.543R, 111R+PBS.873F. The fragments synthesized via PCR were sequenced and assembled into a contiguous sequence.

25 The results from all these "race" PCR and sequencing reactions led to the assembly of a contiguous fragment of 4559 base pairs. This sequence contains an open reading frame of 2824 base pairs encoding a polypeptide of 907 amino acids as set forth in SEQ ID NO:2 and FIG. 2. Two primers, HG38-3419 and HG38-99FS, were designed and synthesized. These two primers were used to amplify the full-length coding region of HG38. The PCR product was cloned into pCR2.1 30 (Invitrogen, San Diego, CA) by TA cloning.

EXAMPLE 2

DNA Analysis

The sequence of the full length HG38 cDNA is provided in FIGS. 1A-1C (SEQ ID NO:1). The amino acid sequence of this receptor is 5 provided in FIG. 2 (SEQ ID NO:2). FASTA searches and phylogenetic analysis were performed using the program GCG (Genetics Computer Group, Madison, Wisconsin, USA). The analysis revealed that HG38 is a member of the G-protein coupled glycoprotein hormone receptor family. Hydropathy analysis was performed using the program Pepplot 10 of GCG (Genetics Computer Group, Madison, Wisconsin, USA) and showed that HG38 has 7 transmembrane domains typical of the rhodopsin family of G-protein coupled receptors, beginning at about amino acid 557 of SEQ ID NO:2. The deduced polypeptide sequence of HG38 contains several sites for cleavage of a signal peptide from the N- 15 terminus of the protein (FIG. 4).

EXAMPLE 3

Analysis of the pattern of expression of HG38

Multi-tissue Northern blot analysis was performed as 20 follows. Ready-to-use human multi-tissue Northern blots were purchased from Clontech (Clontech, Palo Alto, CA, USA). A total of six blots were used to analyze the expression of HG38 in human tissues.

Random Priming

25 Fragments of the HG38 cDNA were labeled with ^{32}P by random priming using the REDIPRIME® labeling kit (Amersham, Inc., Chicago, IL, USA). Reactions were carried using the protocol of the kit supplier. Approximately 50 ng of DNA in 45 μl of H_2O was boiled for 3 minutes., and then quickly chilled to 0°C for 5 minutes. The DNA 30 solution was transferred to a REDIPRIME® tube and mixed with the lyophilized reagents in the tube. Then, 5.0 μl of a- ^{32}P -dCTP (~5000 Ci/mM) was added and the tube was incubated at 37°C for 15 minutes.

The reaction was stopped by adding 5.0 μ l of 0.5 M EDTA (pH8.0). Unincorporated nucleotides were removed by gel-filtration using a spun column.

5 Northern Hybridization.

The labeled fragments were used as probes for HG38 RNA. Hybridizations were carried out in the ExpressHyb buffer of Clontech following the protocol provided by the membrane supplier Clontech (Palo Alto, CA, USA). The membranes were prehybridized at 68°C for 1 hr in 10 the Expresshyb buffer with gentle agitation. The 32 P-labeled probe was denatured by adding NaOH to a final concentration of 0.2 N and then added into the hybridization solution. Hybridizations were performed for 3 hours at 68°C. The membranes were removed from the hybridization buffer and washed once in 2x SSC, 0.1% SDS, for 10 min. at room 15 temperature. The membranes were then washed at 0.1xSSC, 0.1% SDS for 30 minutes at 50°C. The blots were analyzed using a Phosphaimager (Molecular Dynamics, Sunnyvale, CA, USA).

Analysis.

20 HG38 transcripts were detected in skeletal muscle, spinal cord, placenta, and various regions of the brain. (FIG. 5) The greatest levels of expression of HG38 were found in skeletal muscle, placenta and spinal cord. More moderate levels of expression were observed in various regions of the brain. In all of these tissues, the major transcript 25 of HG38 is ~5.0 kb. A minor transcript of ~3.0 kb was detected together with the major transcript.

EXAMPLE 4

Isolation of genomic DNA encoding HG38

30 The HG38 cDNA is used as a probe to isolate human genomic DNA encoding the receptor. The cDNA can be used in its entirety or portions of the sequence can be used. If portions of the sequence less than 100 nucleotides are used as a probe, one should perform homology

analysis of the selected probe sequence against human sequences in general to assess the uniqueness of the chosen sequence in human DNA. If the chosen sequence exhibits high homology to a variety of human DNA sequences, then that sequence will not perform well as a probe specific for HG38 genomic DNA. For example, portions of the cDNA encoding amino acid sequences that are highly conserved among G-protein coupled receptors can be used. However, in that case one should expect to identify receptor genes in addition to HG38, and a large number of identified fragments should be studied further. Thereafter, one will be required to determine which of the identified DNAs encodes HG38. This can be accomplished simply by sequencing the identified genomic DNA fragments and comparing the sequences to HG38 sequence provided herein (SEQ ID NO:1).

Once a probe sequence has been selected the probe is labeled by any means known in the art, including but not limited to incorporation of radioisotopes or biotin. Under appropriately stringent conditions, the probe is hybridized against a library of human genomic DNA fragments. The stringency of the hybridization reaction can be adjusted by means known in the art, *e.g.*, varying salt concentrations and temperature, to obtain appropriately specific hybridization of the probe to the target sequence. The fragments identified by the probe can be sequenced or subjected to restriction enzyme digestion to confirm that they contain HG38 genomic DNA.

It is possible that the entire genomic gene may not be contained within any one identified fragment. In that case, one will be required to perform chromosome walking, *e.g.*, using an identified fragment as a probe to isolate additional fragments that overlap in the chromosome, to isolate the entire gene. If the isolation of overlapping fragments is required, one can use known methods of manipulation of DNA to construct a contiguous DNA fragment encoding the entire HG38 genomic DNA.

EXAMPLE 5

Transgenic animals

Transgenic animals expressing HG38 as a transgene are provided as follows. A polynucleotide having an HG38 nucleotide sequence, *e.g.*, the nucleotide sequence of a cDNA or genomic DNA encoding a full length HG38 receptor, or a polynucleotide encoding a partial sequence of the receptor, sequences flanking the coding sequence, or both, can be combined into a vector for the integration of the polynucleotide into the genome of an animal. The HG38 sequence can be from a human HG38 or from the animal's HG38.

In this example, the target cell for transgene introduction is a murine embryonic stem cell (ES). ES cells can be obtained from pre-implantation embryos of a variety of non-human animals cultured *in vitro* and fused with embryos (M. J. Evans *et al.*, *Nature* 292:154-156 (1981); Bradley *et al.*, *Nature* 309:255-258 (1984); Gossler *et al.* *Proc. Natl. Acad. Sci. USA* 83:9065-9069 (1986); and Robertson *et al.*, *Nature* 322:445-448 (1986)).

The transgene is introduced into the murine ES cells by microinjection, however, a variety of standard techniques such as DNA transfection, or retrovirus-mediated transduction can be used. The injected ES cells are then combined with blastocysts from a non-human animal. The introduced ES cells colonize the embryo and contribute to the germ line of the resulting chimeric animal (R. Jaenisch, *Science* 240: 1468-1474 (1988)). The chimeric mice are screened for individuals in which germline transformation has occurred. These are crossed to produce animals homozygous for the transgene.

The targeted recombination events as well as the resulting mice are evaluated by techniques well known in the art, including but not limited to DNA (Southern) hybridization to detect the targeted allele, polymerase chain reaction (PCR), polyacrylamide gel electrophoresis (PAGE) and Western blots to detect DNA, RNA and protein.

Three basis types of transgenic animals are created depending on the construction of the transgene vector. If the vector is designed to include a nucleotide sequence that encodes a full length

human HG38 receptor and to integrate at a site other than the animal's endogenous HG38 gene, the resultant transgenic animal will express both a native and human HG38 receptors. If the vector is designed without a cognate HG38 gene and to integrate at the site of the animal's 5 endogenous HG38 gene such that after integration the endogenous gene is altered to such an extent that the animal lacks a functional HG38 receptor, then a knockout animal is produced. Finally, if the vector is designed to replace the endogenous HG38 gene with a human gene, or is designed to change the sequence of the endogenous gene to encode the 10 amino acid sequence of the human gene, *i.e.*, is humanized, then the resultant animal lacks a native HG38 receptor and expresses a human HG38 receptor. Animals having a human gene and lacking an endogenous gene can also be created by crossing the first type of animal with a knockout animal to obtain animals homozygous for the knockout 15 and homozygous for the added human HG38 gene. This can be facilitated if the human gene integrates in a chromosome different from the chromosome carrying the endogenous HG38 gene.

Transgenic animals are a source of cells and tissues for use in assays of HG38 modulation, activation or inhibition. Cells can be 20 removed from the animals, established as cell lines and maintained in culture as convenient.

EXAMPLE 6

Assay for ligands of the HG38 receptor

25 Glutathione S-transferase ("GST") HG38 receptor fusion constructs.

Polypeptide fusion constructs are made by inframe fusion of all or a portion of the N-terminal ligand-binding domain of the HG38 G-protein coupled glycoprotein hormone receptor and the carboxy 30 terminus of the GST gene. The disclosure of SEQ ID NOS:1-2 allow the artisan of ordinary skill to construct any such nucleic acid molecule encoding a GST-HG38 fusion protein. In particular, fusions can be constructed using a polynucleotide that encodes the N-terminal

fragment of HG38 from amino acids about 22 to about 557, or 22 to the end of the sequence of SEQ ID NO:2, fused to GST C-terminus.

Soluble recombinant HG38 fusion proteins can be expressed in various expression systems, some of which are described herein, 5 including *Spodoptera frugiperda* (Sf21) insect cells using a baculovirus expression vector (e.g., Bac-N-Blue DNA from Invitrogen or pAcG2T from Pharmingen).

The fusion protein is then loaded onto a glutathione column. The C-terminal domain of GST binds to the glutathione and the 10 N-terminal region of HG38 is exposed to the buffer phase. After washing the column, a sample that may contain a ligand of the HG38 receptor is passed over the column. The sample can be cell or tissue extracts, bodily fluids or compounds or molecules that are purified or synthesized. The sample can be applied directly or after dilution or dialysis in a buffer 15 approximating physiological conditions. Ligands of the receptor are bound by the N-terminal domain of HG38. After washing the column the ligands are eluted. This can be achieved, for example, by applying a gradient of NaCl to the column in wash buffer. Unknown ligands present in biological extracts or fluids can be characterized by standard 20 chemical and biochemical methods. Ligands identified in this method can be used as candidates in assays for agonists or antagonists of the HG38 receptor.

Assays for ligands can also be conducted as described below for assays for agonist and antagonists of HG38. A candidate compound 25 or molecule that shows agonist or antagonist activity can also be a ligand for HG38.

EXAMPLE 7

Assays for agonists and antagonists of the receptor

30 In any assay using recombinant host cells it is first necessary to produce the cells as described elsewhere herein. Briefly, a polynucleotide of the present invention is used to transform or transfect the appropriate cells, or cells can be obtained and cultured from an appropriate transgenic animal.

Melanophore system.

The melanophore screening system is described in WO 92/01810, published February 6, 1992. Briefly, melanophores are transfected to express the HG38 G-protein coupled receptor. In an assay for antagonists, the transformed melanophores are exposed to both an activating ligand and a candidate compound. Inhibition of the signal generated by the ligand indicates that the candidate is a potential antagonist of the receptor. In an assay for an agonist, the cells are contacted with candidate compounds and it is determined whether any compound activates the receptor to generate a signal. Activation of the receptor indicates that the candidate is a potential agonist of the receptor.

Yeast expressing mammalian adenylate cyclase.

Screening methods employing yeast that express mammalian adenylate cyclase are described in WO 95/30012, published November 9, 1995. These yeast can be engineered to co-express the HG38 receptor in the presence of an appropriate G-protein. In an assay for antagonists, the transformed yeast are exposed to both an activating ligand of HG38 and a candidate compound. Inhibition of the signal generated by the ligand indicates that the candidate is a potential antagonist of the receptor. In an assay for an agonist, the cells are contacted with candidate compounds and it is determined whether any compound activates the receptor to generate a signal. Activation of the receptor indicates that the candidate is a potential agonist of the receptor.

Yeast pheromone protein surrogate screening.

Yeast cells engineered to produce pheromone system protein surrogates can be used to screen for the ability of the surrogate to substitute for the cognate yeast pheromone receptor as described in WO 94/23025, published October 13, 1994. Generally, the method involves expressing the HG38 G-protein coupled receptor in *Saccharomyces cerevisiae* in which the receptor is linked to pheromone pathway. In this system, the yeast Ga subunit is generally deleted and replaced with a mammalian Ga protein so that the mammalian G protein-coupled receptor can be coupled to the yeast pheromone pathway. Members of a

plasmid library capable of expressing peptides of random sequences are introduced into an appropriate yeast strain. Clones encoding agonist ligands for the HG38 receptor can be selected for their stimulation of the pheromone pathway. Clones encoding antagonist ligands for the HG38 receptor can be selected for their inhibition of the pheromone pathway in the presence of an HG38 agonist. Alternatively, libraries of chemicals can be screened for their agonist or antagonist activity by testing the chemicals directly.

10 **Phospholipase second signal screening**

Another screening technique involves expressing the HG38 receptor wherein the receptor is linked to a phospholipase C or D. Cells including CHO, endothelial, embryonic kidney and other cells can be used. As in other screens, ligand and candidates are screened for agonist or antagonist activities by detecting the activation or inhibition or the receptor's activation of the phospholipase second signal. An example of one such system using yeast cells expressing a heterologous phospholipase is found in WO 96/40939, published December 19, 1996.

20 **Yeast two-hybrid system**

The yeast two-hybrid system expressing the HG38 G-protein coupled receptor can be used for screening for agonists and antagonists of the receptor (Fields and Song, 1989, *Nature* 340:245-246). In particular, the entire or portions of the extracellular domain of the G-protein coupled receptor can be fused to the DNA binding domain of transcription factor Gal4 or LexA. Yeast cells expressing these constructs are used to carry out screening for molecules that interact with the G-protein coupled receptor by using standard protocols such as those described previously (Fields and Song, 1989) of the two-hybrid screening method. Such molecules represent potential agonists or antagonists of the receptor.

EXAMPLE 8

Assay for modulators of the receptor

Compounds or molecules that are modulators of the receptor can be detected in assay described or as follows. An antibody specific for the extracellular domain of the receptor is obtained by standard techniques. The antibody can be polyclonal or monoclonal. The affinity of the antibody for the extracellular domain of the receptor should preferably be at least 10^6 , and more preferably at least 10^8 , to simplify conducting the assay. A cell culture that expresses the receptor is provided. The cell culture can be one that naturally expresses the receptor, a cell line stably or transiently transfected with an expression vector including the receptor gene, or derived from a transgenic animal having a transgene including the receptor gene.

Two samples of the culture are used in the assay. One sample is used as a control and is treated with a placebo, *i.e.*, a compound or molecule determined to have no modulatory effects on the receptor in the assay. The second sample is treated with a candidate modulator. At various times after or during treatment a portion of the culture can be withdrawn. The antibody can then be used to qualify or quantify the amount of receptor present on the surface of the cell. This can be done by numerous techniques known in the art including using antibody detectably labeled with ^{125}I , gold, enzyme or other known labels. Alternatively, a detectable label can be carried on a second antibody specific for the first. The amount of receptor found on the cells treated with a potential modulator is quantitatively or qualitatively compared to the amount of receptor found on the control cells. A change in the former relative to the latter is indicative of the whether or not the test compound is a modulator of the receptor.

In an alternative form of the assay one can treat cells as described herein and then isolate the receptors present in treated and control cells. The receptor preparations can be made as crude cell extracts, membrane or intracellular fractions of the cells or after purification steps, *e.g.*, chromatography, precipitation or affinity isolation steps. Crude, partially or highly purified preparations of

receptors can be analyzed for receptor content, *e.g.*, by using antibodies specific for the receptor.

5 In any assay it can be advantageous to devise an internal control so that the results of different runs of assays can be compared to each other. A cellular protein that is unrelated to the receptor and present in relatively constant amounts in the cells used in the assay can serve as an internal control.

EXAMPLE 9

10 Assay for determining whether a compound can bind to the receptor

The present invention includes methods of identifying compounds that specifically bind to an HG38 protein, as well as compounds identified by such methods. The specificity of binding of compounds having affinity for an HG38 protein is shown by measuring the affinity of the 15 compounds for recombinant cells expressing the cloned receptor or for membranes from these cells. Expression of the cloned receptor and screening for compounds that bind to an HG38 protein or that inhibit the binding of a known, radiolabeled ligand of HG38 to these cells, or membranes prepared from these cells, provides an effective method for the rapid selection of 20 compounds with high affinity for an HG38 protein. Such ligands need not necessarily be radiolabeled but can also be nonisotopic compounds that can be used to displace bound radiolabeled compounds or that can be used as activators in functional assays. Compounds identified by the herein method are likely to be agonists or antagonists of HG38 and may be peptides, proteins, 25 or non-proteinaceous organic molecules.

Therefore, the present invention includes assays by which HG38 agonists and antagonists may be identified. Methods for identifying agonists and antagonists of other receptors are well known in the art and can be adapted to identify agonists and antagonists of HG38. Accordingly, the 30 present invention includes a method for determining whether a candidate compound is a potential agonist or antagonist of HG38 that comprises:

(a) transfecting cells with an expression vector encoding an HG38 protein;

(b) allowing the transfected cells to grow for a time sufficient to allow the HG38 protein to be expressed;

(c) exposing the cells to a labeled ligand of an HG38 protein in the presence and in the absence of the candidate compound;

5 (d) measuring the binding of the labeled ligand to the HG38 protein; where if the amount of binding of the labeled ligand is less in the presence of the candidate compound than in the absence of the candidate compound, then the candidate compound is a potential agonist or antagonist of an HG38 protein.

10 The conditions under which step (c) of the method is practiced are conditions that are typically used in the art for the study of protein-ligand interactions: *e.g.*, physiological pH; salt conditions such as those represented by such commonly used buffers as PBS or in tissue culture media; a temperature of about 4°C to about 55°C.

15 The present invention also includes a method for determining whether a candidate compound is capable of binding to an HG38 protein, *i.e.*, whether the candidate compound is a potential agonist or an antagonist of an HG38 protein, where the method comprises:

20 (a) providing test cells by transfecting cells with an expression vector that directs the expression of an HG38 protein in the cells;

(b) exposing the test cells to the candidate compound;

(c) measuring the amount of binding of the candidate compound to the HG38 protein;

25 (d) comparing the amount of binding of the candidate compound to the HG38 protein in the test cells with the amount of binding of the candidate compound to control cells that have not been transfected with an HG38 protein;

30 wherein if the amount of binding of the candidate compound is greater in the test cells as compared to the control cells, the candidate compound is capable of binding to an HG38 protein. Determining whether the candidate compound is actually an agonist or antagonist can then be accomplished by the use of functional assays such as, *e.g.*, the assay involving the use of promiscuous G-proteins described herein.

35 The conditions under which step (b) of the method is practiced are conditions that are typically used in the art for the study of protein-ligand interactions: *e.g.*, physiological pH; salt conditions such as those represented

by such commonly used buffers as PBS or in tissue culture media; a temperature of about 4°C to about 55°C.

In a particular embodiment of the herein-described methods, the cells are eukaryotic cells. In another embodiment, the cells are mammalian cells. In other embodiments, the cells are L cells L-M(TK⁻) (ATCC CCL 1.3), L cells L-M⁻ (ATCC CCL 1.2), 293 (ATCC CRL 1573), Raji (ATCC CCL 86), CV-1 (ATCC CCL 70), COS-1 (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C127I (ATCC CRL 1616), BS-C-1 (ATCC CCL 26) or MRC-5 (ATCC CCL 171).

The assays described herein can be carried out with cells that have been transiently or stably transfected with an HG38 protein. Transfection is meant to include any method known in the art for introducing an HG38 protein into the test cells. For example, transfection includes calcium 15 phosphate or calcium chloride mediated transfection, lipofection, infection with a retroviral construct containing an HG38 protein, and electroporation.

Where binding of the candidate compound or agonist to HG38 is measured, such binding can be measured by employing a labeled candidate compound or agonist. The candidate compound or agonist can be labeled in 20 any convenient manner known to the art, *e.g.*, radioactively, fluorescently, enzymatically.

In particular embodiments of the herein-described methods, the HG38 protein has an amino acid sequence of SEQ ID NO:2.

The herein-described methods can be modified in that, rather 25 than exposing the test cells to the candidate compound, membranes can be prepared from the test cells and those membranes can be exposed to the candidate compound. Such a modification utilizing membranes rather than cells is well known in the art and is described in, *e.g.*, Hess *et al.*, 1992, Biochem. Biophys. Res. Comm. 184:260-268.

Accordingly, the present invention provides a method for 30 determining whether a candidate compound is capable of binding to an HG38 protein comprising:

- (a) providing test cells by transfecting cells with an expression vector that directs the expression of an HG38 protein in the cells;
- 35 (b) preparing membranes containing the HG38 protein from the test cells and exposing the membranes to a ligand of an HG38 protein

under conditions such that the ligand binds to the HG38 protein in the membranes;

(c) subsequently or concurrently to step (b), exposing the membranes from the test cells to a candidate compound;

5 (d) measuring the amount of binding of the ligand to the HG38 protein in the membranes in the presence and the absence of the candidate compound;

10 (e) comparing the amount of binding of the ligand to an HG38 protein in the membranes in the presence and the absence of the candidate compound where a decrease in the amount of binding of the ligand to an HG38 protein in the membranes in the presence of the candidate compound indicates that the candidate compound is capable of binding to an HG38 protein;

15 The present invention provides a method for determining whether a candidate compound is capable of binding to an HG38 protein comprising:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of an HG38 protein in the cells;

20 (b) preparing membranes containing the HG38 protein from the test cells and exposing the membranes from the test cells to the candidate compound;

(c) measuring the amount of binding of the candidate compound to the HG38 protein in the membranes from the test cells;

25 (d) comparing the amount of binding of the candidate compound to the HG38 protein in the membranes from the test cells with the amount of binding of the candidate compound to membranes from control cells that have not been transfected with an HG38 protein;

30 where if the amount of binding of the candidate compound to the HG38 protein in the membranes from the test cells is greater than the amount of binding of the candidate compound to the membranes from the control cells, then the candidate compound is capable of binding to an HG38 protein

EXAMPLE 10

Use of HG38 sequence for gene therapy

5 Nucleic acid according to the present invention, *e.g.* encoding the authentic biologically active HG38 polypeptide or a functional fragment thereof, can be used in a method of gene therapy, to treat a patient who is unable to synthesize the active polypeptide or unable to synthesize it at the normal level, thereby providing the effect provided by the wild-type with the aim of treating and/or preventing one or more symptoms of one or more other diseases.

10 Vectors such as viral vectors have been used to introduce genes into a wide variety of different target cells. Typically the vectors are exposed to the target cells so that transfection can take place in a sufficient proportion of the cells to provide a useful therapeutic or prophylactic effect from the expression of the desired polypeptide. The 15 transfected nucleic acid can be permanently incorporated into the genome of each of the targeted cells, providing long lasting effect, or alternatively the treatment may have to be repeated periodically.

20 A variety of vectors, both viral vectors and plasmid vectors, are known in the art, *see e.g.* US Patent No. 5,252,479 and WO 93/07282. In particular, a number of viruses have been used as gene transfer vectors, including adenovirus, papovaviruses, such as SV40, vaccinia virus, herpesviruses, including HSV and EBV, and retroviruses, including gibbon ape leukemia virus, Rous Sarcoma Virus, Venezuelan equine encephalitis virus, Moloney murine leukemia 25 virus and murine mammary tumorvirus. Many gene therapy protocols have used disabled murine retroviruses.

30 Disabled virus vectors are produced in helper cell lines in which genes required for production of infectious viral particles are expressed. Helper cell lines are generally missing a sequence which is recognised by the mechanism which packages the viral genome and produce virions which contain no nucleic acid. A viral vector which contains an intact packaging signal along with the gene or other sequence to be delivered (*e.g.* encoding the HG38 polypeptide or a

fragment thereof) can be packaged in the helper cells into infectious virion particles, which can then be used for the gene delivery.

Other known methods of introducing nucleic acid into cells include electroporation, calcium phosphate co-precipitation, mechanical 5 techniques such as microinjection, transfer mediated by liposomes and direct DNA uptake and receptor-mediated DNA transfer. Liposomes can encapsulate RNA, DNA and virions for delivery to cells. Depending on factors such as pH, ionic strength and divalent cations being present, the composition of liposomes can be tailored for targeting of particular 10 cells or tissues. Liposomes include phospholipids and may include lipids and steroids and the composition of each such component can be altered. Targeting of liposomes can also be achieved using a specific binding pair member such as an antibody or binding fragment thereof, a protein, a sugar or a glycolipid.

15 The aim of gene therapy using nucleic acid encoding the polypeptide, or an active portion thereof, is to increase the amount of the expression product of the nucleic acid in cells in which the level of the wild-type polypeptide is absent or present only at reduced levels. Such treatment can be therapeutic or prophylactic, particularly in the 20 treatment of individuals known through screening or testing to have an HG38 allele associated with a disease state and hence a predisposition to the disease.

Similar techniques can be used for anti-sense regulation of gene expression, e.g. targeting an antisense nucleic acid molecule to 25 cells in which a mutant form of the gene is expressed, the aim being to reduce production of the mutant gene product. Other approaches to specific down-regulation of genes are well known, including the use of ribozymes designed to cleave specific nucleic acid sequences. Ribozymes are nucleic acid molecules, actually RNA, which specifically cleave 30 single-stranded RNA, such as mRNA, at defined sequences, and their specificity can be engineered. Hammerhead ribozymes can be preferred because they recognize base sequences of about 11-18 bases in length, and so have greater specificity than ribozymes of the *Tetrahymena* type 35 which recognise sequences of about 4 bases in length, though the latter type of ribozymes can also be useful in certain circumstances as will be recognized by one of skill in the art. References on the use of ribozymes

include Marschall, et al. 1994. *Cellular and Molecular Neurobiology* 14(5):523; Hasselhoff, 1988. *Nature* 334:585 and Cech, 1988. *J. Amer. Med. Assn.* 260:3030.

5

EXAMPLE 11

Construction of polynucleotides encoding an HG38 receptor protein

The full length amino acid sequence of the HG38 receptor protein is provided in SEQ ID NO:2. A native human cDNA, sequence including an open reading frame encoding the amino acid sequence of HG38, is provided in SEQ ID NO:1. Because of the degeneracy of the genetic code, the sequence of the open reading frame provided in SEQ ID NO:1 is only one of many nucleotide sequences that can encode the amino acid sequence of HG38. One of ordinary skill in the art is familiar with the genetic code and can, using standard techniques of molecular biology, can generate polynucleotides having alternative nucleotide sequences that encode the same amino acid sequence provided in SEQ ID NO:2.

Alternative nucleotide sequences can be DNA, RNA, mixtures of DNA and RNA or can include alternative linkages between nucleotides as described herein.

WHAT IS CLAIMED:

1. An isolated polynucleotide selected from the group consisting of:
 - (a) a polynucleotide encoding a polypeptide having an amino acid sequence of SEQ ID NO:2.
 - (b) a polynucleotide which is complementary to the polynucleotide of (a),
 - (c) a polynucleotide representing a polymorphic form of (a), and
 - (d) a polynucleotide comprising at least 20 nucleotides of the polynucleotide of (a), (b) or (c), said 20 nucleotides being highly specific for an HG38 gene.
2. The polynucleotide of claim 1 wherein the polynucleotide comprises nucleotides selected from the group consisting of natural, non-natural and modified nucleotides.
3. The polynucleotide of claim 1 wherein the internucleotide linkages are selected from the group consisting of natural and non-natural linkages.
4. The polynucleotide of claim 1 that includes the entire nucleotide sequence of SEQ ID NO:1.
- 25 5. The polynucleotide of claim 1 that includes at least a nucleotide sequence of the open reading frame of SEQ ID NO:1.
- 30 6. The polynucleotide of claim 5 having a sequence of human genomic DNA.
7. The polynucleotide of claim 5 having a sequence of a human RNA.
- 35 8. The polynucleotide of claim 1 that includes a nucleotide sequence that encodes a polypeptide having the amino acid sequence from about 22 to about 557 of SEQ ID NO:2.

9. An expression vector comprising a polynucleotide of
claim 1.

5 10. A host cell comprising the expression vector of
claim 9.

11. A process for expressing a HG38 receptor protein in a
recombinant host cell, comprising:

10 (a) introducing an expression vector of claim 9 into a
suitable host cell; and,

(b) culturing the host cells of step (a) under conditions
which allow expression of said the HG38 protein from said expression
vector.

15 12. A substantially purified polypeptide having an amino
acid sequence selected from the group consisting of

(a) a polypeptide having an amino acid sequence of SEQ
ID NO:2,

20 (b) a polypeptide having at least an amino acid sequence
from about amino acid 22 to about 557 of SEQ ID NO:2,

(c) a polypeptide having at least an amino acid sequence
from about amino acid 22 to about the end of SEQ ID NO:2, and

(d) a polypeptide representing a polymorphic form of (a),
25 (b) or (c).

13. A method of determining whether candidate
compounds or molecules are agonists of a polypeptide of claim 12
comprising:

30 (a) providing a cell expressing on the surface thereof a
polypeptide of claim 12, said polypeptide being associated with second
component which provides a detectable signal when an agonist binds to
the polypeptide,

35 (b) contacting said cell with the compound or molecule
under conditions sufficient to permit the binding of the candidate, and

(c) determining whether the candidate is an agonist by detecting a signal produced by said second component.

14. The method of claim 13 wherein said second component is a G-protein and said detectable signal is a cleavage of GTP by said G-protein.

15. A method of determining whether candidate compounds or molecules are antagonists of a polypeptide of claim 12 comprising:

(a) providing a cell expressing on the surface thereof a polypeptide of claim 12, said polypeptide being associated with second component which provides a detectable signal when an antagonist binds to the polypeptide,

15 (b) contacting said cell with the compound or molecule under conditions sufficient to permit the binding of the candidate, and

(c) determining whether the candidate is an antagonist by detecting a signal produced by said second component.

20 16. The method of claim 15 wherein said second component is a G-protein and said detectable signal is a failure of said G-protein to cleave GTP.

25 17. A method of isolating a polynucleotide encoding a mutant or polymorphic form of an HG38 receptor comprising:

(a) providing a detectable probe highly specific for HG38 polynucleotides,

(b) providing a sample containing polynucleotides having human sequences;

30 (c) contacting said probe with said sample under conditions sufficient for the hybridization of said probe with a polynucleotide encoding a mutant or polymorphic form of an HG38 receptor,

(d) detecting any probe-polynucleotide hybrids, and

35 (e) isolating the polynucleotide encoding the mutant or polymorphic form of an HG38 receptor.

18. A transgenic mouse comprising a transgene having a polynucleotide of claim 1.

5 19. The mouse of claim 18 further comprising a knockout of the endogenous murine HG38 receptor gene.

10 20. The mouse of claim 19 wherein said murine HG38 receptor gene is humanized.

15 21. An antibody specific for a mammalian HG38 receptor.

22. A method for determining whether a candidate compound 15 is capable of binding to an HG38 protein comprising:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of an HG38 protein in the cells;

(b) exposing the test cells to the candidate compound ;

(c) measuring the amount of binding of the candidate

20 compound to the HG38 protein;

(d) determining whether a candidate compound is capable of binding to an HG38 protein by comparing the amount of binding of the candidate compound to the HG38 protein in the test cells with the amount of binding of the candidate compound to control cells that have not been

25 transfected with an HG38 protein.

23. A method for determining whether a candidate compound is capable of binding to an HG38 protein comprising:

30 (a) providing test cells by transfecting cells with an expression vector that directs the expression of an HG38 protein in the cells;

(b) preparing membranes containing the HG38 protein from the test cells and exposing the membranes from the test cells to the candidate compound ;

35 (c) measuring the amount of binding of the candidate compound to the HG38 protein in the membranes from the test cells;

(d) determining whether a candidate compound is capable of binding to the HG38 protein by comparing the amount of binding of the candidate compound to the HG38 protein in the membranes from the test cells with the amount of binding of the candidate compound to membranes from 5 control cells that have not been transfected with an HG38 protein.

1/11

1 CGCACCGCCA CTGTCGCCGC TGCAGCCAGG CCTGCTCCGA AGGCCCCCGT
51 GGCAGCAACC GGCACCTAG TCCCCGCCGC GCTTCTCCCTC GCGGCCACG
101 CCGTGGGGTC AGGAACGCGG CGTCTGGCGC TGCAGACGCC CGCTGAGTTG
151 CAGAAGCCCA CGGAGCGGGC CCCGGCCGCC CACGGCCCGT AGCAGTCCGG
201 TGCTGCTCTC CGCCCGCGTC CGGCTCTGG CCCCTACTT CGGGCACCAT
251 GGACACCTCC CGGCTCGGTG TGCTCCTGTC CTTGCCGTG CTGCTGCAGC
301 TGGCAGCCGG GGGCAGCTCT CCCAGGCTG GTGTGTTGCT GAGGGGCTGC
351 CCCACACACT GTCATTGCGA GCCCCACGGC AGGATGTTGC TCAGGGTGGA
401 CTGCTCCGAC CTGGGGCTCT CGGAGCTGCC TTCCAACCTC AGCGTCTTC
451 CCTCCTACCT AGACCTAGT ATGAACAAACA TCAGTCAGCT GCTCCCGAAT
501 CCCCTGCCCA GTCTCCGCTT CCTGGAGGAG TTACGTCTG CGGGAAACGC
551 TCTGACATAC ATTCCCAAGG GACCATTCAC TGGCCTTAC AGTCTTAAAG
601 TTCTTATGCT GCAGAATAAT CAGCTAAGAC ACGTACCCAC AGAAGCTCTG
651 CAGAATTGCA GAAGCCTCA ATCCCTGGT CTGGATGCTA ACCACATCAG
701 CTATGTGCC CCAAGCTGTT TCAGTGGCT GCATCCCTG AGGCACCTGT
751 GGCTGGATGA CAATGCGTTA ACAGAAATCC CCGTCCAGGC TTTAGAAGT
801 TTATCGGCAT TGCAAGCCAT GACCTTGGCC CTGAACAAAA TACACCACAT
851 ACCAGACTAT CCCTTTGGAA ACCTCTCCAG CTTGGTAGTT CTACATCTCC
901 ATAACAATAG AATCCACTCC CTGGAAAGA AATGCTTGA TGGGCTCCAC
951 AGCCTAGAGA CTTTAGATT AAATTACAAT AACCTTGATG AATTCCCCAC
1001 TCCAATTAGG ACACTCTCCA ACCTTAAAGA ACTAGGATT CATAGCAACA
1051 ATATCAGGTC GATACTCGAG AAACCATTT TAGCCAACCC TTCTTTATT
1101 ACAATACATT TCTATGACAA TCCCACCAA TTTGGGGAA GATCTGCTTT
1151 TCAACATTAA CCTGAACTAA GAACACTGAC TCTGAATGGT CCCTCACAAA
1201 TAACTGAATT TCCCTGATTAA ACTGGAACCTG CAAACCTGGA GACTCTGACT
1251 TTAACCTGGAG CACAGATCTC ATCTCTTCC CAAACCGTCT GCAATCAGTT
1301 ACCTAATCTC CAAGTGCCTAG ATCTGTCTTA CAACCTATTAA GAAGATTTAC
1351 CCAGTTTTTC AGTCTGCCAA AAGCTTCAGA AAATTGACCT AAGACATAAT
1401 GAAATCTACG AAATTAAAGT TGACACTTTC CAGCAGTTGC TTAGCCTCCG
1451 ATCGCTGAAT TTGGCTGGAA ACAAAATTGC TATTATTAC CCAATGCAT
1501 TTTCACCTTT GGCATCCCTA ATAAAGCTGG ACCTATCGTC CAAACCTCTG
1551 TCGCTTTTC CTATAACTGG GTTACATGGT TTAACACTACT TAAAATTAAAC
1601 AGGAAATCAT GCCTTACAGA GCTTGATATC ATCTGAAAAC TTTCCAGAAC
1651 TCAAGCTTAT AGAAATGCCT TATGCTTACG AGTGCCTGTGTT ATTGGAGTG
1701 TGTGAGAATG CCTATAAGAT TTCTAATCAA TGGAAATAAAC GTGACAACAC
1751 CAGTATGGAC GACCTTCATA AGAAAGATGC TGGAAATGTTT CAGGCTCAAC
1801 ATGAACCTGA CCTGAAAGAT TTCCCTGCTTG ACTTTGAGGA AGACCTGAAA
1851 CCCCTTCATT CAGTGCAGTG TTCACTTCC CCAGGCCCT TCAAACCCCTG
1901 TGAACACCTG CTTGATGGCT GGCTGATCAG AATTGGAGTG TGGACCATAAC
1951 CAGTCTGGC ACTTACTTGT AATGCTTGG TGACTTCAAC AGTTTCAGA
2001 TCCCCCTCTGT ACATTCCCC CATTAAACTG TTAATTGGGG TCATCGCACC
2051 AGTGAACATG CTCACGGGAG TCTCCAGTGC CGTGCCTGGCT GGTGTGGATG

FIG. 1A

2/11

2101 CGTTCACTT TGGCAGCTT GCACGACATG GTGCCCTGGT GGAGAATGG
2151 GTTGGTTGCC ATGTCATTGG TTTTTGTC ATTGTTGCTT CAGAAATCATC
2201 TGTTTCCTG CTTACTCTGG CAGCCCTGGA CGCTGGGTT TCTGTGAAAT
2251 ATTCTGAAAC ATTGAAACG AAAGCTCCAT TTTCTAGCCT GAAAGTAATC
2301 ATTTGCTCT GTGCCCTGCT GGCCTTGACC ATGCCCGAG TTCCCCCTGCT
2351 CGGTGGCAGC AAGTATGGCG CCTCCCCCTCT CTGCCTGCCT TTGCTTTTG
2401 GGGAGCCCGAG CACCATGGGC TACATGGTCG CTCTCATCTT GCTCAATTCC
2451 CTTGCTTCC TCATGATGAC CATTGCCTAC ACCAAGCTCT ACTGCAATT
2501 GGACAAGGGA CACCTGGAGA ATATTTGGGA CTGCTCTATG GTAAAACACA
2551 TTGCCCTGTT GCTCTTCACC AACTGCATCC TAAACTGCCG TGTGGCTTTC
2601 TTGTCCTTCT CCTCTTTAAT AACCTTACA TTTATCAGTC CTGAAGTAAT
2651 TAAGTTTATC CTTCTGGTGG TAGTCCCCTCT TCCCTGCATGT CTCATCCCC
2701 TTCTCTACAT CTGTTCAAT CCTCACTTA AGGAGGATCT GGTGAGCCTG
2751 AGAAAGCAAA CCTACGTCTG GACAAGATCA AAACACCCAA GCTTGATGTC
2801 AATTAACCTCT GATGATGTGG AAAAACACTC CTGTGACTCA ACTCAAGCCT
2851 TGGTAACCTT TACCAGCTCC AGCATCACTT ATGACCTGCC TCCCAGTTCC
2901 GTGCCATCAC CAGCTTATCC ACTGACTGAG AGCTGCCATC TTTCCTCTGT
2951 GGCATTGTC CCATGCTCT AATTAATATG TGAAGGAAAA TGTTTCAAA
3001 CGTTGAGAAC CTGAAAATGT GAGATTGACT ATATCAGAGC AGTAATTAAAT
3051 AAGAAGAGCT GAGGTGAAAC TCGGTTAAA AACCAAAAAA GAATCTCTCA
3101 GTTAGTAAGA AAAGGCTGAA AACCTTGTG TACTTGAGAG TGAATATAAG
3151 TCTAAATGCT GCTTTGTATA ATTTGTTAG CTAACCCATA CATCGATCAC
3201 ACTATTAAAG TGAGCCCAGA TCAAAAAAGC AGATTGAAAT TTCTTTAGA
3251 AAAGATTCTC CATGATTGAA ATTGCATTCT CTTAAACTC ACCAATGAA
3301 TCATTTGGG AGGTGGGAGA ACCCCCTGT TTTCCAATG CGTTTATTIA
3351 AACCCACAAA CTCAGAGGT TGTTGGGGA ATTAGGAAAA TAAGGGTTT
3401 CAATGACCTA CATTGCTAGG TAGAGGCTGT GATCCATGGG TTTCATTCTA
3451 ATGACCATGT GAGATGTTG GTCTTCCCTT GCTTCCCTCA GAAAGATCCT
3501 TCTAAGGCAC AAATCCCTA GATGGATAAT GTAAAGTATT GTTAACCTCAC
3551 TCATATTGAG ATCATTTTA GAGATACCG GTTTATGTA TCAGCACTAG
3601 ATGGTCCAC CCTCATGGGA TAAAAGCT TACAAGTATT TTGAAAGAAA
3651 AACTGACCAA AATTCTAAA TTGTTACTAA GGCAATCATG CACAGGTGAC
3701 GTATGCTTA TCTGATTGT TTGTTACTCC TTGGTGGCCA AAGCTCAGAA
3751 GGGAAATTCCA CTGCCAGCAA TGAACATACC TCGAAAAGAA AGTAAGCAAT
3801 CTGGGATTT TTTCTGGGT TAGTAAAGAA TTGTTCAAT AAGTTTATC
3851 AGTTGATTCA AACTGATGTG CATCTTAATG ATCAAATGTG CACATTACAT
3901 AAATTAAGTC CACTGATACA ACTTCTTACA CATGTATCTC TAGTAGCTCT
3951 GGCAAAACCCAA ATATCTGACA CCACTTTGGA CTCAAGAGAC TCAGTAACGT
4001 ATTATCCTGT TTATTTAGCT TGTTTACG TGTGTTCTCT CTGGATAACC
4051 CACTTGATGT TAGGAACATT ATTTCTCTGC TTATTCATA TTAATACTGT

FIG. 1B

3/11

4101 GTTACGTATT TTAAGAACCA AGTTATTAAA TAAGAAAAGT CAAAGTATT
4151 ATTCTTACCT TCTATTATCC TATATTAGCT TCAATACATC CAAACCAAAT
4201 GGCTGTTAGG TAGATTATT TTTATATAAG CATGTTTATT TTGATCAGAT
4251 GTTTAACCTT GGATTTGAAA AAATACATT ATGAGATGTT TTATAAGATG
4301 TGTAAATATA G MCTGTATT TATTACTATA GTAAAGGTTC AGTAACATTA
4351 AGGACCATGA TAATGATAAT AAACCTTGTAA CAGTGGCATA TTCTTGATT
4401 TATATTGTGT TTCTCTGCCCTT AATTCTTTA AATTCTTAA CTGTATATAT
4451 GTAAATATAT AGTACTTGTAA AATAGATTCC AAATTGCTT TTCTATTGGG
4501 TAAAAAATAA ATTGTATAA AAATGTGTGA CTATGAAACA AAAAAAAA
4551 AAAAAAAA

FIG. 1C

4/11

1 MDTSLGVLL SLPVLLQLAT GGSSPRSGVL LRGCPTHCHC EPDGRMILLRV
51 DCSDLGLSEL PSNLSVFTSY LDLSMNNISQ LLPNPLPSLR FLEELRLAGN
101 ALTYIPKGAF TGLYSLKVLM LQNNQLRHVP TEALQNLRSI QSLRLDANHI
151 SYVPPSCFSG LHSRLHLWLD DNALTEIPVQ AFRSLSALQA MTLALNKH
201 IPDYAFGNLS SLVVLHLHNN RIHSLGKKCF DGLHSLETLD LNYNNLDEFP
251 TAIRTLSNLK ELCFHSNNIR SIPEKAFVGN PSLITIHFYD NPIQFVGRSA
301 FQHLPELRTL TLNGASQITE FPDLTGTLN ESLTLTGAQI SSLPQTVCNQ
351 LPNLQVLDLS YNLLEDLPSF SVCQKLQKID LRHNEIYEIK VDTFQQLLSL
401 RSLNLAWNKI AIIHPNAFST LPSLIKLDLS SNLLSSFPIT GLHGLTHLKL
451 TGNHALQSLI SSENFPELKV IEMPTYAYQCC AFGVCENAYK ISNQWNKGDN
501 SSMDDLHKKD AGMFQAQDER DLEDFLLDPE EDLKALHSVQ CSPSPGPFKP
551 CEHLLDGWL I RICVWTIAVL ALTCNALVTS TVFRSPLYIS PIKLLIGVIA
601 AVNMLTGSS AVLAGVDAFT FGSFARHGAW WENVGCHVI GFLSIFASES
651 SVFLLTLAAL ERGFSVKYSA KFETKAPFSS LKVIIILCAL LALTMAAVPL
701 LGGSKYGASP LCLPLPFGEP STMGYMVALI LLNSLCFLMM TIAYTKLYCN
751 LDKGDLENIW DCSMVKHIAL LLFTNCILNC PVAFLSFSSL INLTFISPEV
801 IKFILLVVP LPACLNPLLY ILFNPHFKED LVSLRKQTYV WTRSKHPSLM
851 SINSDDVEKQ SCDSTQALVT FTSSSITYDL PPSVPSPAY PVTESCHLSS
901 VAFVPC

FIG.2

5/11

10	30	50
CACCGCCACTGTCGCCGCTGCAGCCAGGGCTGCTCCGAAGGCCGGCTGGCGGCAACCGC		
70	90	110
CACCTCAGTCCCCGGCGCTTCTCTCGCCGCCACGCCGTGGGTCAAGAACGCCG		
130	150	170
TCTGGCGCTGCAGACGCCCGCTGAGTTGCAGAACCCACGGAGCGGCCGGCGCCA		
190	210	230
CGGCCCCTAGCAGTCCGTGCTGCTCTCCGCCCGCTCCGGCTCGTGGCCCCCTACTTCG		
250	270	290
GGCACCATGGACACCTCCGGCTGGTGTGCTCCTGTCCTGGCTGCTGCTGCAGCTG		
Met Asp Thr Ser Arg Leu Gly Val Leu Leu Ser Leu Pro Val Leu Leu Gln Leu		
310	330	350
GCGACCGGGGGCAGCTCTCCAGGTCTGGTGTGCTGAGGGGCTGCCAACACACTGT		
Ala Thr Gln Gly Ser Ser Pro Arg Ser Gln Val Leu Leu Arg Gly Cys Pro Thr His Cys		
370	390	410
CATTCCGACCCGACGGCAGGATGTTGCTCAGGGTGGACTGCTCCGACCTGGGCTCTCG		
His Cys Gln Pro Asp Gln Arg Met Leu Leu Arg Val Asp Cys Ser Asp Leu Gln Leu Ser		
430	450	470
GAGCTGCCTTCCAACCTCAGCGTCTTACCTCACCTACCTAGACCTCAGTATGAACAAACATC		
Gln Leu Pro Ser Asn Leu Ser Val Phe Thr Ser Tyr Leu Asp Leu Ser Met Asn Asn Ile		
490	510	530
ACTCAGCTGCTCCGAATCCCCTGCCAGTCTCCGCTCTGGAGGACTTACCTCTTGCCT		
Ser Gln Leu Leu Pro Asn Pro Leu Pro Ser Leu Arg Phe Leu Gln Gln Leu Arg Ala		
550	570	590
GGAAACGCTCTGACATACTTCCAAGGGAGCATTCACTGGCCTTACAGTCTTAAAGTT		
Gly Asn Ala Leu Thr Tyr Ile Pro Lys Gln Ala Phe Thr Gln Leu Tyr Ser Leu Lys Val		
610	630	650
CTTATGCTGAGAATAATCAGCTAACACAGTACCCACAGAACGCTCTCCAGAATTGCGA		
Leu Met Leu Gln Asn Asn Gln Leu Arg His Val Pro Thr Gln Ala Leu Gln Asn Leu Arg		

FIG. 3A

6/11

670	690	710
AGCCTCAATCCCTCGCTCGATGCTAACACATCAGCTATGTGCCCCAAGCTGTTTC		
SerLeuGlnSerLeuArgLeuAspAlaAsnHisIleSerTyrValProProSerCysPhe		
730	750	770
AGTGGCCTGCATTCCCTGAGGCACCTGTGGCTGGATGACAATGCGTTAACAGAAATCCCC		
SerGlyLeuHisSerLeuArgHisLeuTrpLeuAspAspAsnAlaLeuThrGluIlePro		
790	810	830
GTCCAGGCTTTAGAAGTTATCGGCATTGCAAGCCATGACCTGGCCCTGAACAAAATA		
ValGlnAlaPheArgGlyLeuSerAlaLeuGlnAlaMetThrLeuAlaLeuAsnIle		
850	870	890
CACCAACATACCAGACTATGCCTTGGAAACCTCTCCAGCTGGTAGTTCTACATCTCCAT		
HisHisIleProAspTyrAlaPheGlyAsnLeuSerSerLeuValValLeuHisLeuHis		
910	930	950
AACAAATAGAACATCCACTCCCTGGGAAAGAAATGCTTGATGGGCTCCACAGCCTAGAGACT		
AsnAsnArgIleHisSerLeuGlyLysLysCysPheAspGlyLeuHisSerLeuGluThr		
970	990	1010
TTAGATTAAATTACAATAACCTTGATGAATTCCCCACTGCAATTAGGACACTCTCCAAC		
LeuAspLeuAsnTyrAsnAsnLeuAspGluPheProThrAlaIleArgThrLeuSerAsn		
1030	1050	1070
CTTAAAGAACTAGGATTTCATACCAACAATATCAGGTCGATACCTGAGAAAGCATTGTA		
LeuLysGluLeuGlyPheHisSerAsnAsnIleArgSerIleProGluLysAlaPheVal		
1090	1110	1130
GGCAACCCTCTCTTATTACAATACATTCTATGACAATCCCCTCAATTGTTGGGAGA		
GlyAsnProSerLeuIleThrIleHisPheTyrAspAsnProIleGlnPheValGlyArg		
1150	1170	1190
TCTGCTTTCAACATTACCTGAACTAAGAACACTGACTCTGAATGGTGCCTCACAAATA		
SerAlaPheGlnHisLeuProGluLeuArgThrLeuThrLeuAsnGlyAlaSerGlnIle		
1210	1230	1250
ACTGAATTCCGATTTAACCTGAACTGCAAACCTGGAGACTCTGACTTTAACGGAGCA		
ThrGluPheProAspLeuThrGlyThrAlaAsnLeuGluSerLeuThrLeuThrGlyAla		

FIG. 3B

7/11

1270	1290	1310
CAGATCTCATCTTCCTCAAACCGCTGCAATCAGTTACCTAATCTCCAAGTGCTAGAT		
GlnIleSerSerLeuProGlnThrValCysAsnGlnLeuProAsnLeuGlnValLeuAsp		
330	350	1370
CTGTCTTACAACCTATTAGAACGATTACCCAGTTTCAGTCTGCCAAAGCTTCAGAAA		
LeuSerTyrAsnLeuLeuGluAspLeuProSerPheSerValCysGlnLysLeuGlnLys		
1390	1410	1430
ATTGACCTAAGACATAATGAAATCTACGAAATTAAAGTTGACACTTCCAGCAGTTGCTT		
IleAspLeuArgHisAsnGlnIleTyrGluIleLysValAspThrPheGlnGlnLeuLeu		
1450	1470	1490
AGCCTCCGATCGCTGAATTGGCTTGAACAAAATTGCTATTATTCACCCCAATGCATT		
SerLeuArgSerLeuAsnLeuAlaTrpAsnLysIleAlaIleIleHisProAsnAlaPhe		
1510	1530	1550
TCCACTTGGCATCCCTAATAAACCTGGACCTATCGTCCAACCTCCTGTCGTCTTCTT		
SerThrLeuProSerLeuIleLysLeuAspLeuSerSerAsnLeuLeuSerSerPhePro		
1570	1590	1610
ATAACTGGTTACATGGTTACTCACTAAACAGGAAATCATGCCTTACAGAGC		
IleThrGlyLeuHisGlyLeuThrHisLeuLysLeuThrGlyAsnHisAlaLeuGlnSer		
1630	1650	1670
TTGATATCATCTGAAACTTCCAGAACTCAAGCTTATAGAAATGCCTTATGCTTACACAGC		
LeuIleSerSerGluAsnPheProGluLeuLysValIleGluMetProTyrAlaTyrGln		
1690	1710	1730
TGCTGTGCATTTGGACTCTGAGAACGCTATAAGAAAGATGCTGGAATGTTCAGGCTAACAGT		
CysCysAlaPheGlyValCysGluAsnAlaTyrLysIleSerAsnGlnTrpAsnLysGly		
1750	1770	1790
GACAACAGCACTATGGACGACCTTCATAAGAAAGATGCTGGAATGTTCAGGCTAACAGAT		
AspAsnSerSerMetAspAspLeuHisLysLysAspAlaGlyMetPheGlnAlaGlnAsp		
1810	1830	1850
GAACGTGACCTTGAACATTCCCTGCTGACTTGGAGGAAGACCTGAAAGCCCTTCATTCA		
GluArgAspLeuGluAspPheLeuLeuAspPheGluGluAspLeuLysAlaLeuHisSer		

FIG. 3C

8/11

1870	1890	1910
CTCCAGTGTACCTCCCCAGGCCCCCTCAAACCTGTAAACACCTGCTTGATGGCTGG		
ValGlnCysSerProSerProGlyProPheLysProCysGluHisLeuLeuAspGlyTrp		
1930	1950	1970
CTGATCAGAATTGGACTGTGGACCATAGCACCTGGCACTTACTGTAATGCTTGGTG		
LeuIleArgIleGlyValTrpThrIleAlaValLeuAlaLeuThrCysAsnAlaLeuVal		
1990	2010	2030
ACTTCAACAGTTTCAGATCCCCTGTACATTCCCCATTAAACTGTTAATTGGGTC		
ThrSerThrValPheArgSerProLeuTyrIleSerProIleLysLeuLeuIleGlyVal		
2050	2070	2090
ATCGCAGCAGTGAACATGCTCACGGGAGTCTCCAGTGCCGTGCTGGCTGGATGCG		
IleAlaAlaValAsnMetLeuThrGlyValSerSerAlaValLeuAlaGlyValAspAla		
2110	2130	2150
TTCACTTTGGCAGTTGCACGACATGGTGCCTGGTGGAGAATGGGGTTGGTGGCAT		
PheThrPheGlySerPheAlaArgHisGlyAlaTrpTrpGluAsnGlyValGlyCysHis		
2170	2190	2210
GTCATTGGTTTTGTCCATTTGCTTCAGAACATCTGTTCTGCTTACTCTGGCA		
ValIleGlyPheLeuSerIlePheAlaSerGluSerSerValPheLeuLeuThrLeuAla		
2230	2250	2270
GCCCTGGAGCGTGGTTCTCTGTGAAATATTCTGAAAATTGAAACGAAAGCTCCATT		
AlaLeuGluArgGlyPheSerValLysTyrSerAlaLysPheGluThrLysAlaProPhe		
2290	2310	2330
TCTAGCCTGAAAGTAATCATTGCTGTGCC~CCTGGCCTGACCATGGCCGAGTT		
SerSerLeuLysValIleIleLeuLeuCysAlaLeuLeuAlaLeuThrMetAlaAlaVal		
2350	2370	2390
CCCTGCTGGTGGCAGCAAGTATGGCGCTCCCCCTCTGCCTGCCTTGCCTTGG		
ProLeuLeuGlyGlySerLysTyrGlyAlaSerProLeuCysLeuProLeuProPheGly		
2410	2430	2450
GAGCCCAGCACCAGGGCTACATGGCTGGCTCATCTGCCTCAATTCCCTTGCTTCCTC		
GluProserThrMetGlyTyrMetAlaAlaLeuIleLeuLeuAsnserLeucyspheLeu		
2470	2490	2510
ATGATGACCATTGCCTACACCAAGCTACTGCAATTGGACAAGGGAGACCTGGAGAAT		
MetMetThrIleAlaTyrThrLysLeuTyrCysAsnLeuAspLysGlyAspLeuGluAsn		

FIG. 3D

9/11

2530	2550	2570
ATTCGGACTGCTCATGGTAAAACACATTGCCCTGTCCTTCACCAACTGCATCTA		
IleTrpAspCysSerMetValLysHisIleAlaLeuLeuLeuPheThrAsnCysIleLeu		
2590	2610	2630
AACTGCCCTGCGTTCTGCTCTCCTCTTAATAAACCTTACATTATCACTCCT		
AsnCysProValAlaPheLeuSerPheSerSerLeuIleAsnLeuThrPheIleSerPro		
2650	2670	2690
GAAGTAATTAAAGTTATCCTCTGGTGGTAGTCCCACCTCCTGCATGTCTCAATCCCCT		
GluValIleLysPheIleLeuLeuValValValProLeuProAlaCysLeuAsnProLeu		
2710	2730	2750
CTCTACATCTGTTCAATCCTCACTTAAAGGAGGATCTGGTAGCCTGAGAAAGCAAACC		
LeuTyrIleLeuPheAsnProHisPheLysGluAspLeuValSerLeuArgLysGlnThr		
2770	2790	2810
TACGTCTGGACAAGATCAAAACACCCAAAGCTTGATGTCATTAACCTCTGATGATGTCGA		
TyrValTrpThrArgSerLysHisProSerLeuMetSerIleAsnSerAspAspValGlu		
2830	2850	2870
AAACAGTCCTGTGACTCAACTCAAGCCTGGTAACCTTACCAAGCTCCAGCATCACTTAT		
LysGlnSerCysAspSerThrGlnAlaLeuValThrPheThrSerSerSerIleThrTyr		
2890	2910	2930
GACCTGCCTCCCAGTTCCGTGCCATACCCAGCTTACCAAGCTGAGAGCTGCCATCTT		
AspLeuProProSerSerValProSerProAlaTyrProValThrGluSerCysHisLeu		
2950	2970	2990
TCCTCTGTCGCAATTGTCCTCATGTCCTAAATTAAATGTAACGAAAATGTTCAAAGG		
SerSerValAlaPheValProCysLeuEnd		
3010	3030	3050
TTGAGAACCTGAAATGTGAGATTGAGATATCAGAGCAGTAATTAAATAAGAACAGCTGA		
3070	3090	3110
CGTAAACTCGTTAAAAACCAAAAAAGAATCTCTCAGTTAGTAACGAAACGCTGAAAA		
3130	3150	3170
CCTCTGATACTTGAGAGTGAATATAAGTCTAAATGCTGTTGTATAATTCTTCAGCT		
3190	3210	3230
AAGGGATAGATCGATCACACTATTAAAGTGAGCCCAGATCAAAAAGCAGATTGAAATT		
3250	3270	3290
TCTTAGAAAGATTCTCCATGATTGAATTGCATTCTCTTAAACTCACCAATGAA		

FIG. 3E

10/11

Score 9.3 at residue 22

Sequence: LLSPVLLQLATG-CSSPRSGVLLRGCPTHCHCEPDGRMLLRVD
| (signal) | (mature polypeptide) |
9 22 51

Score 5.1 at residue 25

Sequence: LPVLLQLATGGSS-PRSGVLLRGCPTHCHCEPDGRMLLRVDCSD
| (signal) | (mature polypeptide) |
12 25 54

Score 4.8 at residue 41

Sequence: GVLLRGCPTHCHC-EPDGRMLLRVDCSDLGLSELPNSLSVFTSY
| (signal) | (mature polypeptide) |
28 41 70

Score 4.5 at residue 26

Sequence: PVLLQLATGGSSP-RSGVLLRGCPTHCHCEPDGRMLLRVDCSDL
| (signal) | (mature polypeptide) |
13 26 55

Score 4.0 at residue 24

Sequence: SLPVLLQLATGGS-SPRSGVLLRGCPTHCHCEPDGRMLLRVDCS
| (signal) | (mature polypeptide) |
11 24 53

Score 3.6 at residue 21

Sequence: VLLSLPVLLQLAT-GCSSPRSGVLLRGCPTHCHCEPDGRMLLRV
| (signal) | (mature polypeptide) |
8 21 50

FIG.4

11/11

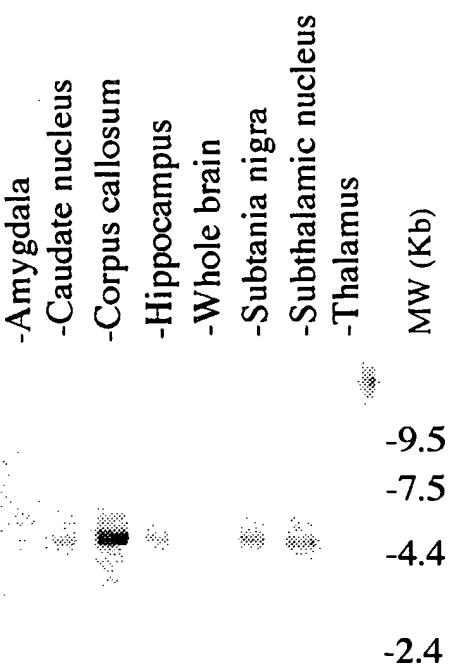


FIG.5C



FIG.5A

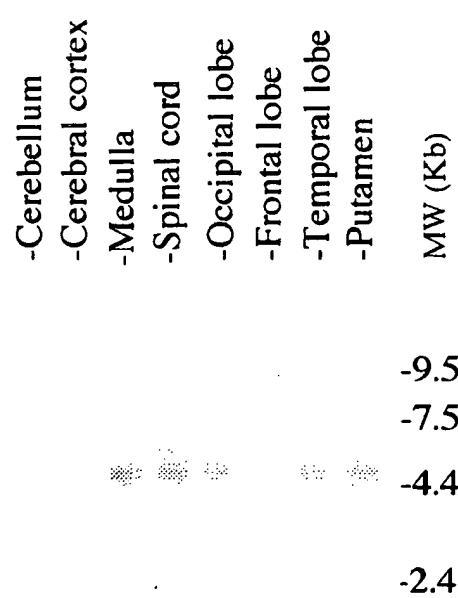


FIG.5D

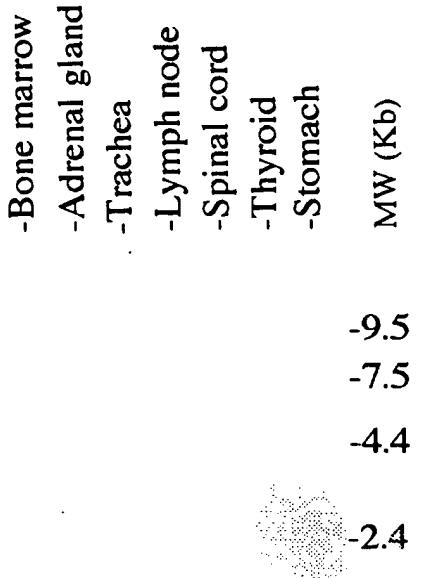


FIG.5B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/19979

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :536/23.5; 435/320.1, 7.1, 6, 7.2; 530/350, 387.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5; 435/320.1, 7.1, 6, 7.2; 530/350, 387.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN, APS, hg38, receptor#, hg(w)38, g(w)protein(w)coupled(w)glycoprotein#, (g(w)protein)(10A)(hormone(w)receptor#)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REICHERT et al., Structure-function relationships of the glycoprotein hormones and their receptors, Trends in Pharmacological Sciences. May 1991, Volume 12, pages 199-203, see entire document.	1-23

Further documents are listed in the continuation of Box C. See patent family annex.

• Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
• A document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
• E earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
• L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Z"	document member of the same patent family
• O document referring to an oral disclosure, use, exhibition or other means		
• P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
11 DECEMBER 1998	12 JAN 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized Officer <i>JAMES MARTINELL</i> JAMES MARTINELL
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/19979

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12N 15/12, 15/63, 15/00; C12P 21/02; C07K 14/47, 16/00; G01N 33/53; C12Q 1/68